ELSEVIER International Journal of Approximate Reasoning 22 (1999) 217–248

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

# Local computation of Gaussian belief functions

Liping Liu *

*Department of Management, Southern Illinois University, Carbondale, IL 62901-4627, USA*

Accepted 1 August 1999

**Abstract**

Gaussian belief functions represent logical and probabilistic knowledge for mixed variables, some of which are deterministic, some vacuous, and some Gaussian. They include as special types linear equations, statistical observations, multivariate Gaussian distributions, and vacuous belief functions. The notion of Gaussian belief functions was proposed by A.P. Dempster (Normal belief functions and the Kalman filter, Technical report, Department of Statistics, Harvard University, Cambridge, MA, 1990.), formalized by G. Shafer (A note on Dempster's Gaussian belief functions, Technical report, School of Business, University of Kansas, Lawrence, KS, 1992.) and L. Liu (International Journal of Approximate Reasoning 14 (1996) 95–126.); (in: D. Fisher, Hans-J. Lenz (Eds.), Learning Models from Data: AI and Statistics V, Springer, New York, NY, 1996, pp. 79–88.) and successfully applied in combining independent statistical models in L. Liu (Model combination using Gaussian belief functions, Technical report, School of Business, University of Kansas, Lawrence, KS, 1995.). In this paper, we propose a join-tree computation scheme for expert systems using Gaussian belief functions. We first represent Dempster's rule of combination obtained in Liu (1996) alternatively in terms of matrix sweepings. We then show the operations of Gaussian belief functions follow the axioms of P.P. Shenoy and G. Shafer (in: R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, vol. 4, North-Holland, Amsterdam, 1990, pp. 169–198.) and justify the possibility of a join-tree computation scheme for Gaussian Belief functions. The result enriches the theory of local computation by extending its applicability to the combination of statistical models and the integration of knowledge bases. Examples are carried out to illustrate how

* Tel. : +1-618-453-7885.
*E-mail address:* liu@cba.siu.edu (L. Liu)


PII: S0888-613X(99)00029-8

combined inference can be made in accordance with multiple statistical models using graphically structured belief function models. 

*Keywords:* Expert systems; Dempster–Shafer theory; Belief networks; Local computation; Gaussian belief functions; Graphical models; Multivariate Gaussian distribution

## 1. Introduction

Local computational algorithms are applicable to Bayesian or belief function expert systems with discrete and finite variables. Problems involving continuous variables can be solved approximately using simulation techniques. However, there are some important exceptions. Lauritzen and Wermuth [11] extend the join-tree approach to Bayesian networks with mixed variables, some of which are discrete and some conditionally Gaussian. Shachter and Kenley [18] extend the reduction algorithm to the assessment and analysis of linear-quadratic-Gaussian influence diagrams. In parallel, this paper proposes and justifies local computation for Gaussian belief functions (GBFs).

The notion of GBFs extends the Dempster–Shafer theory of belief functions in representing mixed knowledge, some of which is logical, some statistical, and some vacuous. Logical knowledge is represented by a hyperplane in the sample space. Ignorance is represented by partitioning a hyperplane into parallel sub-hyperplanes as focal elements. Statistical knowledge is represented by a Gaussian distribution across the focal elements over the hyperplane. In its full generality, a GBF may be used to represent a wide range of statistical and knowledge-based models, notably linear models and Gaussian belief networks. It includes as special cases non-probabilistic linear equations, statistical observations, multivariate Gaussian distributions, and vacuous belief functions.

GBFs have a wide range of real applications. Dempster [4,5] shows how the Kalman filter can be understood in terms of GBFs. As Dempster shows, the state equations and the observation equations are deterministic knowledge, which can be captured by logical belief functions. The distributional assumptions on independent random disturbances are probabilistic knowledge, which can be represented by Bayesian belief functions. Finally, the observations are represented as another set of logical belief functions. The recursion involved in the filter can be regarded as a special case of the recursion involved in the computation of Gaussian belief function marginals. The full Kalman filter model results from judging all these component belief functions to be independent, and combining them into a single belief function according to Dempster's rule.

Liu [12] applies the theory of GBFs to the integration of independent knowledge bases and statistical models. In details, Liu [12] treats different sources of data as independent items of evidence and represents the discovered knowledge (e.g., linear models or Bayesian belief networks) by GBFs. Then one can combine independent models in a fashion we combine GBFs and make enriched inferences based on the combined model. As Liu [12] shows, the GBF method can combine models of different kinds that may involve different variables. Therefore, it generalizes the meta-analysis for integrating independent statistical findings [2,6] by lifting the restriction that the models to be combined be the same and the parameters to be estimated be common. Nevertheless, in the restricted case, the GBF method has similar flavor to the meta-analysis. For example, for the problem of weighted mean [30] and its generalization [2], both methods give the same posterior distribution of common regression coefficients.

Dempster [5] sketches how local computation works for belief functions in general. However, it is an open question whether it works for GBFs [21]. The existing work in this area [9,24,27,28] applies to finite and to condensable belief functions, but not to GBFs, which are usually continuous but not condensable. Presumably the justification for the finite case can be extended to a justification for the continuous case by a straightforward limiting argument, but this has not been done to date.

The primary purpose of this paper is to justify join-tree computation for GBFs by proving that combination and marginalization of GBFs follow the axioms of Shenoy and Shafer [28]. We also propose a join-tree algorithm for GBFs by adapting the Shafer–Shenoy architecture for propagation. We show how to construct a join-tree for GBFs and how to propagate messages and compute marginals in a join-tree. We carry out a comprehensive example to illustrate the join-tree computation algorithm for making combined predictions and inferences based on multiple statistical models.

An outline of the rest of this paper is as follows. To motivate the reader, in Section 2, we examine two definitions of GBFs and their dual relationship. In Section 3, we define two generalized matrix sweeping operations and re-express Dempster's rule obtained in Liu [14] in terms of the sweepings. In Section 4, we prove the axioms of Shenoy and Shafer [28] and propose a join-tree algorithm for GBFs. Finally, we conclude the paper in Section 5.

## 2. Gaussian belief functions

This section reviews two notions of GBFs and establishes their dual relationship. It attempts to describe a GBF in its full generality but in less technical terms. Except Section 2.3, much of this section elaborates on Shafer [21] and Liu [14].

### 2.1. Gaussian belief functions in a sample space

Let $\mathbf{V}^*$ be an $n$-dimensional sample space. A belief function on $\mathbf{V}^*$ is defined in general by a basic probability assignment over a class of focal elements [19].

A GBF is special in the sense that its focal elements are the members of a partition of a hyperplane, which are parallel sub-hyperplanes, and its basic probability assignment is a Gaussian distribution across the sub-hyperplanes. A GBF represents our belief regarding the location of the true value in $\mathbf{V}^*$ as follows. We are certain that the true value is on a hyperplane but we do not know its exact location. Along some dimensions of the hyperplane, we believe the true value could be anywhere from $-\infty$ to $+\infty$ and the probability of being at a particular location is described by a Gaussian distribution. Along other dimensions, our knowledge is vacuous, i.e., we believe the location of the true value is somewhere from $-\infty$ to $+\infty$ but the associated probability is unknown. An example of GBFs is Bayesian linear models, whose linear equations define a hyperplane in the sample space. Among the variables linked by the equations, some take on a value with certainty such as observables, some are assumed to be Gaussian such as 'residuals', and some may bear vacuous knowledge such as 'effects', which is usually represented by an improper prior in the Bayesian theory.

Formally, let $\mathbf{C}^*$ denote the hyperplane on which we are sure that the true value lies. We call $\mathbf{C}^*$ *the certainty hyperplane*. We partition $\mathbf{C}^*$ into a family of parallel sub-hyperplanes, as focal elements. Each sub-hyperplane extends from $-\infty$ to $+\infty$ along the dimensions on which our belief is vacuous. Over the cross section of the sub-hyperplanes there is a Gaussian distribution. The value of the distribution function on each sub-hyperplane defines the basic probability for the focal element. All the focal elements constitute a continuum and cannot be enumerated by a finite list as we do for a finite belief function. However, since they are parallel and disjoint, each point $t$ on $\mathbf{C}^*$ marks a unique focal element. Thus, we can use $t$ and $\mathbf{B}^*$ to represent a typical focal element, where $\mathbf{B}^*$ is a sub-hyperplane on $\mathbf{C}^*$ passing through $t$. We call *$t$ the mark of a GBF*. Since all other focal elements are parallel to $\mathbf{B}^*$, they are determined implicitly if $\mathbf{B}^*$ is specified.

In sum, three symbols, $\mathbf{C}^*$, $t$, and $\mathbf{B}^*$ can designate all the focal elements for a GBF. $\mathbf{C}^*$ specifies where they are and $t$ and $\mathbf{B}^*$ specify an example of them. $\mathbf{C}^*$ and $\mathbf{B}^*$ are nested hyperplanes in $\mathbf{V}^*$: $\mathbf{B}^* \subset \mathbf{C}^* \subset \mathbf{V}^*$. Each focal element like $\mathbf{B}^*$ represents a possible region of discernment. The basic probability in terms of 'densities' assigned to it represents how likely it contains the true value. As in a usual belief function, our total belief in terms of the 'densities' allocated to a focal element does not necessitate the re-allocation of any partial belief to its subsets. Thus, each focal element like $\mathbf{B}^*$ is a *no-opinion expressed hyperplane*. A GBF on $\mathbf{V}^*$ can be seen as a Gaussian distribution across the focal elements that are parallel to $\mathbf{B}^*$. In the special case when $\mathbf{B}^*$ is 0-dimensional, the GBF is simply a multivariate Gaussian distribution over $\mathbf{C}^*$. In general, it can be imagined as a normal distribution over a cross section of ellipsoidal cylinders formed by the focal elements. The cylinders are concentric around the central focal element, denoted by $\mathbf{E}^*$, which specifies the expected region of discernment. We call $\mathbf{E}^*$ *the expectation hyperplane*. Of course, $\mathbf{E}^*$ is parallel to $\mathbf{B}^*$ and corresponds to the highest 'density' value. The periphery of each ellipsoidal cylinder is comprised of the focal elements that have the same 'density' value.

As commonly known, a Gaussian distribution is determined by its quadratic log 'density' function. When the domain of the distribution is the whole sample space, its mean vector and covariance matrix can specify the quadratic function. However, when the distribution is defined on a hyperplane, we have to use an inner product on the hyperplane as a generalized quadratic function. Similarly, for GBFs, we can specify the shape, scale, and direction of the Gaussian distribution by a wide-sense inner product $\pi^*(x,y)$ on $\mathbf{C}^*$ with an arbitrary focal element such as $\mathbf{B}^*$ as its null hyperplane, i.e., $\pi^*(x,x) = 0$ iff $x \in \mathbf{B}^*$. Note that basic probabilities are assigned to sub-hyperplanes rather than individual points of $\mathbf{C}^*$. Therefore, we expect that $\pi^*(x,y)$ is invariant if $x$ or $y$ varies within a sub-hyperplane parallel to $\mathbf{B}^*$. Liu [14] shows that the invariance is equivalent to the imposition $\pi^*(x,x) = 0$ iff $x \in \mathbf{B}^*$. This result implies the difference of the representation of a regular Gaussian distribution from that of a GBF. The former is represented by a regular inner product, which is null at a single point, while the latter is represented by a wide-sense inner product, which is null at a hyperplane.

For any fixed $x^0 \in \mathbf{C}^*$, $\pi^*(x^0,x)$ is a linear functional on $\mathbf{C}^*$ with null hyperplane $\mathbf{B}^*$. In addition, $\pi^*(x^0,x)$ is invariant iff $x^0$ is in a hyperplane parallel to $\mathbf{B}^*$, i.e., for each hyperplane $\mathbf{H}^*$ that is parallel to $\mathbf{B}^*$, $H^*(x) = \pi^*(x^0,x)$ ($x^0 \in \mathbf{H}^*$) is a linear functional on $\mathbf{C}^*$ that is zero on $\mathbf{B}^*$ and the choice of $x^0$ does not matter. On the other hand, by slightly modifying the Riesz representation theorem [17], Liu [14] shows that, for any linear functional $H^*(x)$ on $\mathbf{C}^*$ that is zero on $\mathbf{B}^*$, there is a unique hyperplane $\mathbf{H}^*$ parallel to $\mathbf{B}^*$ such that $H^*(x) = \pi^*(x^0,x)$ ($x^0 \in \mathbf{H}^*$). Therefore, linear functional that are zero on $\mathbf{B}^*$ and hyperplanes that are parallel to $\mathbf{B}^*$ are in a one-to-one correspondence. Note that $\mathbf{E}^*$ is parallel to $\mathbf{B}^*$. As a corollary, hyperplane $\mathbf{E}^*$ and linear functional $E^*(x) = \pi^*(x^0,x)$ ($x^0 \in \mathbf{E}^*$) are one-to-one correspondent. Therefore, we can use $E^*(x)$ to specify the center of a Gaussian distribution across the hyperplanes that are parallel to $\mathbf{B}^*$.

Therefore, we arrive at the representation ($\mathbf{C}^*$, $t$, $\mathbf{B}^*$, $\pi^*$, $E^*$) for a GBF. We write $t$ before $\mathbf{B}^*$, $\pi^*$, and $E^*$ because all these objects depend on the choice of $t$. In sum $(\mathbf{C}^*, t, \mathbf{B}^*, \pi^*, E^*)$ expresses beliefs about which element of $\mathbf{V}^*$ is the true value. We are certain that the true value is on the certainty hyperplane $\mathbf{C}^*$. Within $\mathbf{C}^*$, our belief is distributed over ellipsoidal cylinders around a smaller dimensional expectation hyperplane $E^*$. The wide sense inner product $\pi^*$ specifies the shape, scale, and direction of the ellipsoidal cylinders, and the linear functional $E^*$ specifies the expectation hyperplane $\mathbf{E}^*$ by giving its inner product with every other hyperplane parallel to $\mathbf{B}^*$ within $\mathbf{C}^*$. By regarding $\mathbf{V}^*$ to be a coordinate-free linear space, $\pi^*$ is a purely geometric operation that does not depend on the choice of a basis for $\mathbf{V}^*$. Therefore, a GBF

$(\mathbf{C}^*, t, \mathbf{B}^*, \pi^*, E^*)$ can be imagined as a geometric object without referring to its mean vector and covariance matrix.

We can appreciate the broad conception of GBFs by considering their non-trivial special cases. Let $n$, $n-c$, and $n-b$ denote the dimension numbers of $\mathbf{V}^*$, $\mathbf{C}^*$, and $\mathbf{B}^*$, respectively. In general, $c \leqslant b \leqslant n$. By appropriately setting up one or two of the dimension numbers $c$, $b$, and $n$, a GBF can be degenerated into six varieties, which provide building blocks for more complex GBFs. If $b = c = 0$, then the GBF is vacuous and has $\mathbf{V}^*$ as its sole focal element. If $0 < c = b < n$, then the GBF is equivalent to specifying $c$ independent linear equations. If $c = b = n$, the true sample point in $\mathbf{V}^*$ is known with certainty, as might occur by direct observation. If $c = 0$ and $b = n$, then the GBF is an ordinary Gaussian probability distribution on $\mathbf{V}^*$. If $c > 0$ and $b = n$, the GBF is a Gaussian probability distribution over $\mathbf{B}^*$. In the latter two cases, the GBF is Bayesian because its focal elements are singletons with zero dimension. Finally, if $0 = c < b < n$, the GBF is a proper belief function which has a Gaussian distribution for some variables and no-opinions for others.

The advantage of regarding a GBF as a geometrical object is the coordinate-free representation of Dempster's rule of combination. Let $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$ be two GBFs: $\mathrm{Bel}_1 = (\mathbf{C}^{*1}, t, \mathbf{B}^{*1}, \pi^{*1}, E^{*1})$, $\mathrm{Bel}_2 = (\mathbf{C}^{*2}, t, \mathbf{B}^{*2}, \pi^{*2}, E^{*2})$, where $t$ is their common mark. Based on Dempster's rule, Liu [14] shows their combination $\mathrm{Bel}_1 \otimes \mathrm{Bel}_2$ is

$$(\mathbf{C}^{*1} \cap \mathbf{C}^{*2}, t, \mathbf{B}^{*1} \cap \mathbf{B}^{*2}, (\pi^{*1} + \pi^{*2})|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}, (E^{*1} + E^{*2})|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}), \tag{1}$$

where $(\pi^{*1} + \pi^{*2})|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}$ and $(E^{*1} + E^{*2})|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}$ are, respectively, the restriction of $\pi^{*1} + \pi^{*2}$ and $E^{*1} + E^{*2}$ on $\mathbf{C}^{*1} \cap \mathbf{C}^{*2}$.

Note that Eq. (1) is intuitive according to Dempster's rule. As we know, each focal element in $\mathrm{Bel}_i$ is a hyperplane on $\mathbf{C}^{*i}$ that is parallel to $\mathbf{B}^{*i}$, $i = 1, 2$. Therefore, by Dempster's rule, $\mathbf{B}^{*1} \cap \mathbf{B}^{*2}$ is a typical focal element for the combined belief function. Its associated basic probability assignment is obtained by multiplying the basic probabilities assigned to $\mathbf{B}^{*1}$ and $\mathbf{B}^{*2}$ by an appropriate normalization constant. Thus, the log 'density' of the combined GBF is the sum of the component log 'densities'. Therefore, $\pi^* = (\pi^{*1} + \pi^{*2})|_{\mathbf{C}^{*1} \cap \mathbf{C}^{*2}}$ is the basic probability assignment for the focal element $\mathbf{B}^{*1} \cap \mathbf{B}^{*2}$ in terms of log 'densities'.

### 2.2. Gaussian belief functions in a variable space

Let $\mathbf{V}$ be a variable space – a finite linear space spanned by the variables of interest. A GBF can be equivalently represented in $\mathbf{V}$ by specifying a linear functional $E$ and a wide-sense inner product $\pi$ to represent the the mean of each variable and the covariance between variables, respectively.

In general, a GBF may carry mixed knowledge, some of which is logical, some uncertain, and still some vacuous. Logical knowledge, which is represented by hyperplane $\mathbf{C}^*$ in $\mathbf{V}^*$, can be represented in $\mathbf{V}$ by designating some variables in $\mathbf{V}$ to be deterministic and take on the expected value with certainty. Let $\mathbf{C}$ denote the space spanned by all the deterministic variables. Then, $X = E(X)$ iff $X$ is in $\mathbf{C}$ and $\pi(X, Y) = 0$ iff $X$ or $Y$ is in $\mathbf{C}$. We call $\mathbf{C}$ *the certainty space*. Uncertain knowledge on a random variable is represented by its expected value and a non-zero covariance with other variables. Let $\mathbf{B}$ denote the space spanned by those variables on which we have either certain or uncertain knowledge. Then, $E$ and $\pi$ represent both probabilistic and deterministic knowledge for any variable in $\mathbf{B}$. We call $\mathbf{B}$ *the belief space*. Of course, by their definitions, $\mathbf{C}$ and $\mathbf{B}$ are nested subspaces of $\mathbf{V}$ such that $\mathbf{C} \subset \mathbf{B} \subset \mathbf{V}$.

We may have no knowledge about some variables in $\mathbf{V}$. We represent such vacuous knowledge by leaving $E$ and $\pi$ undefined. Therefore, $E$ and $\pi$ are both functionals defined on the belief space $\mathbf{B}$. We are ignorant about the world outside $\mathbf{B}$.

Formally, a GBF on $\mathbf{V}$ is a quadruplet ($\mathbf{C}$, $\mathbf{B}$, $\pi$, $E$). $\pi$ is a wide sense inner product on $\mathbf{B}$ with $\mathbf{C}$ as its null space and $E$ is a linear functional on $\mathbf{B}$. We call $\pi$ the covariance, and $E$ the expectation. $\pi$ and $E$ define a Gaussian distribution for the variables in $\mathbf{B}$ by specifying their means and covariances. This Gaussian distribution is regarded as a full expression of our beliefs, based on a given body of evidence; this item of evidence justifies no beliefs about variables in $\mathbf{V}$ going beyond what is implied by the beliefs about the variables in $\mathbf{B}$. (The evidence might justify some further beliefs about variables that are not in $\mathbf{V}$, but these are outside the conversation so far as a belief function with space $\mathbf{V}$ is concerned.) The Gaussian distribution assigns zero variance to the variables in $\mathbf{C}$; if $X$ is in $\mathbf{C}$, we are certain that it takes the value $E(X)$ with certainty. The GBF contains no knowledge about variables that are not in $\mathbf{B}$. This is represented by restricting the definition of $E$ and $\pi$ on $\mathbf{B}$.

The representation of GBFs in a variable space has an advantage that it renders the definition of marginalization tractable. The marginalization of a GBF in terms of the variable space representation is simply a projection. Suppose $(\mathbf{C}, \mathbf{B}, \pi, E)$ is a GBF, and $\mathbf{M}$ is a subspace of $\mathbf{V}$. Then the marginal of $(\mathbf{C}, \mathbf{B}, \pi, E)$ on $\mathbf{M}$, denoted by $(\mathbf{C}, \mathbf{B}, \pi, E)^{\downarrow\mathbf{M}}$, is a GBF obtained by intersecting certainty space $\mathbf{C}$ and belief space $\mathbf{B}$ with $\mathbf{M}$ and restricting the covariance and the expectation to the new belief space:

$$(\mathbf{C}, \mathbf{B}, \pi, E)^{\downarrow\mathbf{M}} = (\mathbf{C} \cap \mathbf{M}, \mathbf{B} \cap \mathbf{M}, \pi|_{\mathbf{B}\cap\mathbf{M}}, E|_{\mathbf{B}\cap\mathbf{M}}). \tag{2}$$

### 2.3. Duality

As a matter of fact, there exists a dual relationship between the two representations. The hyperplanes $\mathbf{C}^*$ and $\mathbf{B}^*$ and the subspaces $\mathbf{C}$ and $\mathbf{B}$ are constructed based on the same body of evidence. A piece of certain knowledge is represented by $\mathbf{C}^*$ in $\mathbf{V}^*$ and by $\mathbf{C}$ in $\mathbf{V}$ equivalently. Suppose in general $\mathbf{V}$ is spanned by the variables $X_1, X_2, \ldots, X_n$. Let the sample space $\mathbf{V}^* = \{(x_1, x_2, \ldots, x_n) | x_i \in \mathbb{R},\ i = 1, 2, \ldots, n\}$. Then, without loss of generality, we can assume that $\mathbf{C}^*$ is determined by a set of $c$ linearly independent equations as follows:

$$\mathbf{C}^* = \{(x_1, x_2, \ldots, x_n) \,|\, a_{i1}x_1 + \cdots + a_{in}x_n = d_i,\ i = 1, 2, \ldots, c\}. \tag{3}$$

According to Section 2.1, the true value for $X_1, X_2, \ldots,$ and $X_n$ is on $\mathbf{C}^*$ with certainty. In $\mathbf{V}$, the same piece of certain knowledge bears on the variables $a_{i1}X_1 + \cdots + a_{in}X_n,\ i = 1, 2, \ldots, c$. Thus, according to its construction, the certainty space $\mathbf{C}$ is a sub-variable space as follows:

$$\mathbf{C} = \left\{ \sum_{i=1}^{c} \alpha_i (a_{i1}X_1 + \cdots + a_{in}X_n) \,|\, \alpha_i \in \mathbb{R},\ i = 1, 2, \ldots, c \right\}. \tag{4}$$

Since $\mathbf{B}^*$ is a sub-hyperplane contained in $\mathbf{C}^*$, without loss of generality, we assume that $\mathbf{B}^*$ is determined by Eq. (3) and additional $(b - c)$ linearly independent equations:

$$\mathbf{B}^* = \{(x_1, x_2, \ldots, x_n) \,|\, a_{i1}x_1 + \cdots + a_{in}x_n = d_i,\ i = 1, 2, \ldots, b\}. \tag{5}$$

By the semantics of $\mathbf{B}^*$, we have certain knowledge about the variables $a_{i1}X_1 + \cdots + a_{in}X_n,\ i = 1, 2, \ldots, c$, and uncertain knowledge about the variables $a_{i1}X_1 + \cdots + a_{in}X_n,\ i = c + 1, c + 2, \ldots, b$. Therefore, according to Section 2.2, $\mathbf{B}$ is a sub-variable space as follows:

$$\mathbf{B} = \left\{ \sum_{i=1}^{b} \alpha_i (a_{i1}X_1 + \cdots + a_{in}X_n) \,|\, \alpha_i \in \mathbb{R},\ i = 1, 2, \ldots, b \right\}. \tag{6}$$

Eqs. (3)–(6) show the constructive relationships between $\mathbf{C}$ and $\mathbf{C}^*$ as well $\mathbf{B}$ and $\mathbf{B}^*$. Now we further study their mathematical relations by considering the sample space $\mathbf{V}^*$ to be the dual space of $\mathbf{V}$ – the space of all the linear functionals on $\mathbf{V}$. A linear functional $v$ on $\mathbf{V}$ is a real-valued function such that

$$v(\alpha X + \beta Y) = \alpha v(X) + \beta v(Y) \tag{7}$$

for any variables $X$ and $Y$ in $\mathbf{V}$ and real numbers $\alpha$ and $\beta$. We can regard $v(X)$ as a sample value taken on by random variable $X$. In particular, suppose $X_1, X_2, \ldots, X_n$ is a basis for $\mathbf{V}$. Let $v(X_i) = x_i$ for $i = 1, 2, \ldots, n$. Then, $(x_1, x_2, \ldots, x_n)$ is a sample point in $\mathbf{V}^*$. On the other hand, according to Eq. (7), we have

$$v(\alpha_1 X_1 + \cdots + \alpha_n X_n) = (x_1, x_2, \ldots, x_n)(\alpha_1, \alpha_2, \ldots, \alpha_n)^{\mathrm{T}} \tag{8}$$

for any variable $\alpha_1 X_1 + \cdots + \alpha_n X_n$ in $\mathbf{V}$. Thus, linear functional $v$ and sample point $(x_1, x_2, \ldots, x_n)$ are one-to-one correspondent. Therefore, we can use a linear functional and a sample point interchangeably and treat $\mathbf{V}^*$ as the dual space of $\mathbf{V}$. As an specific example, the mark $t = (t_1, t_2, \ldots, t_n)$, which is a sample point in $\mathbf{V}^*$ according to Section 2.1, can be regarded as the linear functional as follows:

$$t(\alpha_1 X_1 + \cdots + \alpha_n X_n) = (t_1, t_2, \ldots, t_n)(\alpha_1, \alpha_2, \ldots, \alpha_n)^{\mathrm{T}} \tag{9}$$

for any $\alpha_1 X_1 + \cdots + \alpha_n X_n$ in $\mathbf{V}$. For another example, the expectation $E$, which is a linear functional that specifies the mean for each variable in $\mathbf{V}$, can be equivalently regarded as the mean vector $\mu$ in $\mathbf{V}^*$.

An advantage with the notion of linear functionals is their independence of the choice of coordinates. In its full generality, $\mathbf{V}$ is a linear vector space abstractly defined by the operation $\alpha X + \beta Y$, where $X$ and $Y$ are variables in $\mathbf{V}$ and $\alpha$ and $\beta$ are real numbers. Accordingly, by Eq. (7), a linear functional is well defined without referring to a basis for $\mathbf{V}$. By regarding $\mathbf{V}^*$ to be the space of linear functionals, $\mathbf{V}^*$ is independent of the basis of $\mathbf{V}$. However, this is not true if we interpret $\mathbf{V}^*$ as the space of sample points. Linear functional $v$ corresponds to sample point $(x_1, x_2, \ldots, x_n)$ in the sense of Eq. (8) if and only if $X_1, X_2, \ldots, X_n$ are chosen to be a basis for $\mathbf{V}$. Had a different basis for $\mathbf{V}$ been chosen, it would correspond to a different sample point.

Regarding $\mathbf{V}^*$ to be the dual space of $\mathbf{V}$ and mark $t$ to be a linear functional on $\mathbf{V}$, we can show that hyperplanes $\mathbf{C}^*$ and $\mathbf{B}^*$ are, respectively, dual to subspaces $\mathbf{C}$ and $\mathbf{B}$ in the sense that

$$\mathbf{C}^* = \{v \mid v(X) = t(X)\ \forall X \in \mathbf{C}\}, \tag{10}$$

$$\mathbf{B}^* = \{v \mid v(X) = t(X)\ \forall X \in \mathbf{B}\}. \tag{11}$$

Eqs. (10) and (11) can be easily verified using Eqs. (3)–(6). Let $v(X_i) = x_i$ for $i = 1, 2, \ldots, n$. Let the mark $t = (t_1, t_2, \ldots, t_n)$. Since $t$ is on $\mathbf{C}^*$, $a_{i1}t_1 + \cdots + a_{in}t_n = d_i$, $i = 1, 2, \ldots, c$. Since $v(X) = t(X)$ for any $X \in \mathbf{C}$, for any $i = 1, 2, \ldots, c$

$$\begin{aligned} v(a_{i1}X_1 + \cdots + a_{in}X_n) &= a_{i1}x_1 + \cdots + a_{in}x_n = t(a_{i1}X_1 + \cdots + a_{in}X_n) \\ &= a_{i1}t_1 + \cdots + a_{in}t_n = d_i. \end{aligned}$$

Therefore, by Eq. (3), Eq. (10) is shown to be true. Similarly we can prove Eq. (11).

Note that, according to Eq. (6), $\mathbf{B}$ does not depend on the constants $d_i$, $i = 1, 2, \ldots, b$, of Eq. (5). Therefore, the construction of the belief space $\mathbf{B}$ does not depend on where we take a typical focal element $\mathbf{B}^*$. For example, if we had chosen $\mathbf{E}^*$ as a focal element, $\mathbf{B}$ would be constructed exactly the same as in Eq. (6). Therefore, the mark $t$ does not enter the representation of a GBF in a variable space. In contrast, according to Eq. (11), the construction of $\mathbf{B}^*$ from $\mathbf{B}$ depends explicitly on the choice of the mark $t$. For example, if we replace $t$ by the expectation $E$ in Eq. (11), then we have

$$\mathbf{E}^* = \{v \mid v(X) = E(X)\ \forall X \in \mathbf{B}\}. \tag{12}$$

Eqs. (3)–(12) jointly imply that the two representation of a GBF can be derived from one another. Actually, Liu [14] takes advantage of this duality and first defines a GBF in **V** and then derives its dual representation in $\mathbf{V}^*$.

Suppose **C**, **B**, and **V** have dimensions $c$, $b$, and $n$, respectively. Then we can choose a basis $X_1, X_2, \ldots, X_n$ of **V** such that $X_1, X_2, \ldots, X_c$ is a basis of **C** and $X_1, X_2, \ldots, X_b$ is a basis of **B**. For $i = 1, 2, \ldots, b$, let $\mu_i$ denote the mean of $X_i$. For $i, j = 1, 2, \ldots, b - c$, let $\Sigma_{ij}$ denote the covariance between $X_{c+i}$ and $X_{c+j}$. Let $\mu = (\mu_1, \mu_2, \ldots, \mu_b)$ and $\Sigma = [\Sigma_{ij}]_{(b-c)\times(b-c)}$. Then, $E$ and $\pi$ can be represented as follows:

$$E[\alpha_1 X_1 + \cdots + \alpha_b X_b] = \mu(\alpha_1, \alpha_2, \ldots, \alpha_b)^{\mathrm{T}}, \tag{13}$$

$$\pi[\alpha_1 X_1 + \cdots + \alpha_b X_b, \beta_1 X_1 + \cdots + \beta_b X_b] = (\alpha_{c+1}, \ldots, \alpha_b)\Sigma(\beta_{c+1}, \ldots, \beta_b)^{\mathrm{T}}, \tag{14}$$

where $\alpha_1 X_1 + \cdots + \alpha_b X_b$ and $\beta_1 X_1 + \cdots + \beta_b X_b$ are any two variables in **B**. It is easy to see that $\pi(\cdot,\cdot)$ is a wide sense inner product on **B** with **C** as its null space: $\pi(X, Y) = 0$ if $X$ or $Y \in \mathbf{C}$.

Let $\mathbf{V}^* = \{(x_1, \ldots, x_n) \,|\, x_i \in \mathbb{R},\ i = 1, 2, \ldots, n\}$ and mark $t = (t_1, \ldots, t_n)$. Then, we have $t_i = \mu_i$ for $i = 1, 2, \ldots, c$ since $t$ is contained in $\mathbf{C}^*$ and each variable in **C** takes on its expected value with certainty. Then, according to Eqs. (10)–(12), we obtain

$$\mathbf{C}^* = \{(x_1, \ldots, x_n) \,|\, x_1 = \mu_1, \ldots, x_c = \mu_c\},$$
$$\mathbf{B}^* = \{(x_1, \ldots, x_n) \,|\, x_1 = \mu_1, \ldots, x_c = \mu_c, x_{c+1} = t_{c+1}, \ldots, x_b = t_b\},$$
$$\mathbf{E}^* = \left\{(x_1, \ldots, x_n) \,|\, x_1 = \mu_1, \ldots, x_c = \mu_c, x_{c+1} = \mu_{c+1}, \ldots, x_b = \mu_b\right\}.$$

As we know, $\pi^*$ is a wide sense inner product defined for all the sub-hyperplanes on $\mathbf{C}^*$ that are parallel to $\mathbf{B}^*$. In the above coordinate system, for any $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbf{C}^*$, $\pi^*(x, y)$ can be represented as a quadratic function:

$$\pi^*(x, y) = (x_{c+1} - t_{c+1}, \ldots, x_b - t_b)\Sigma^{-1}(y_{c+1} - t_{c+1}, \ldots, y_b - t_b)^{\mathrm{T}}.$$

$\mathbf{E}^*$, as per Section 2.1, can be simply written as follows:

$$E^*(x) = (\mu_{c+1} - t_{c+1}, \ldots, \mu_b - t_b)\Sigma^{-1}(x_{c+1} - t_{c+1}, \ldots, x_b - t_b)^{\mathrm{T}}.$$

To represent a GBF in its full generality, advanced notions such as linear functionals and linear spaces are required. However, a general GBF can be often seen as the combination of its special cases, whose individual representation is trivial and can be written as quadratic and linear functions.

## 3. Combination in terms of matrix sweepings

According to Dempster's rule, the combination of two belief functions results from intersecting focal elements and multiplying their associated basic probabilities. Liu [14] adopts this general procedure and defines the combination of GBFs. First, with an appropriate choice of a basis for **V** such that $E$ and $\pi$ are respectively written as a linear and a quadratic functions, a GBF can be seen as a Bayesian belief function for the basis of **V**. From this perspective, Liu [14] defines the combination for any continuous Bayesian belief functions and then derives the combination rule for GBFs as a special case. Furthermore, Liu [14] shows that the resulting rule of combination is equivalent to Eq. (1).

Note that combination and marginalization are respectively better defined in a sample space and a variable space, as shown by Eqs. (1) and (2). To prove the axioms of Shenoy and Shafer [28], we need to interplay them. The combination rule obtained in [14] does not serve this purpose because its representation is implicit. In this section we provide a third equivalent representation of combination in terms of sweep operations. This representation reduces the combination of GBFs into tabular manipulations.

According to Liu [14], in a variable space, the combination of GBFs is determined by conditional means, regression coefficients, and conditional as well as residual covariance matrices. Therefore, the key to the sweep representation is to represent these conditional statistics using sweeping operators, including the forward sweep, denoted by $\rhd$, and the reverse sweep, denoted by $\lhd$. Both $\rhd$ and $\lhd$ were initially proposed to sweep positive definite matrices on certain row and column indices [3]. When applied to a covariance matrix, $\rhd$ implies moving towards a more conditional representation while $\lhd$ means moving to a more marginal representation. Later Dempster [5] adapts the sweep operations to extended matrices. The modified sweep operators are still operated on row and column indices of a matrix. As we will see shortly, the modified sweepings are not sufficient to represent general conditional means and regression coefficients. In this section, we further modify the sweep operators such that they can sweep a matrix on a certain value of certain variables. It will be clear that the operators defined in [5] can be seen as a special case of ours when the given value is zero. It is also interesting to note that many tedious details of regression and distributional analysis can be concisely represented by applying our sweep operators.

Let $[X_1, X_2, \ldots, X_n]$ be a partition of all Gaussian variables with $E(X_i) = \mu_i$ and $\mathrm{Cov}(X_i, X_j) = \Sigma_{ij}$, $i, j = 1, 2, \ldots, n$. The shorthand notation:

$$M = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}_{1-n} \tag{15}$$

called *an extended matrix*, can be used to represent the distribution of the random variables. Note that in Eq. (15), $i$ means rows, $j$ columns, and $1-n$ indicates the partition has indices $1, 2, \ldots, n$. The extended matrix has $(n+1)$ rows, where the first row represents the mean vector and the remaining $n$ rows the covariance matrix. Also, note that the same matrix $M$ can be written differently in the shorthand notation depending on how the variables are partitioned. We define four operations on extended matrices as follows:

**Definition 3.1** (*Marginalization*). Suppose $M$ is an extended matrix for a set of random variables $X$. The marginalization of $M$ to $Y$, denoted by $M^{\downarrow Y}$, is the sub-matrix produced by retaining the rows and columns that corresponds to the variables in $Y$ while deleting the rest.

For example, for the matrix shown in Eq. (15), its marginal to the variables consisting $X_1$ and $X_3$ is as follows:

$$M^{\downarrow [X_1,X_3]} = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}_{1,3}.$$

Note that $Y$ need not be a subset of $X$ in Definition 3.1. Actually, if $Y$ is not a subset of $X$, $M^{\downarrow Y}$ is the same as $M^{\downarrow Y\cap X}$. Obviously, $M^{\downarrow Y}$ represents the marginal distribution for the variables in $Y \cap X$. Since marginalization is simply a restriction, it is easy to show the following consonance rule:

$$(M^{\downarrow Y})^{\downarrow Z} = M^{\downarrow Y \cap Z}. \tag{16}$$

**Definition 3.2** (*Direct sum*). Suppose $M_1$ and $M_2$ are two extended matrices, respectively for $X$ and $Y$. To define their direct sum $M = M_1 \oplus M_2$, we extend $M_1$ to the set $X \cup Y$ by adding zeros to the elements corresponding to the variables that are not in $X$. Similarly, we extend $M_2$ to the set $X \cup Y$ by adding zeros to the elements corresponding to the variables that are not in $Y$. Then $M$ is simply the sum of the extended $M_1$ and $M_2$. It is easy to show the following distributivity rule:

$$(M_1 \oplus M_2)^{\downarrow Y} = (M_1)^{\downarrow Y} \oplus (M_2)^{\downarrow Y}. \tag{17}$$

**Definition 3.3** (*Forward sweep*). Suppose $M$ is an extended matrix for the partition $[X_1, X_2, \ldots, X_n]$. The forward sweep of $M$ on $X_k = x_k$, denoted by $\triangleright(X_k = x_k)M$, is defined as follows:

$$\triangleright(X_k = x_k)\begin{bmatrix} \mu_j \\ \Sigma_{ij} \end{bmatrix}_{1-n} = \begin{bmatrix} \mu_{j.k} \\ \Sigma_{ij.k} \end{bmatrix}_{1-n},$$

where

$$\mu_{j.k} = \begin{cases} \mu_j - (\mu_k - x_k)(\Sigma_{kk})^{-1}\Sigma_{kj}, & j \neq k, \\ \mu_k(\Sigma_{kk})^{-1}, & j = k, \end{cases}$$

$$\Sigma_{ij.k} = \begin{cases} -(\Sigma_{kk})^{-1}, & i = j = k, \\ \Sigma_{ik}(\Sigma_{kk})^{-1}, & j = k \neq i, \\ (\Sigma_{kk})^{-1}\Sigma_{kj}, & i = k \neq j, \\ \Sigma_{ij} - \Sigma_{ik}(\Sigma_{kk})^{-1}\Sigma_{kj}, & \text{otherwise.} \end{cases}$$

**Definition 3.4** (*Reverse sweep*). Suppose $M$ is an extended matrix for the partition $[X_1, X_2, \ldots, X_n]$. The reverse sweep of $M$ on $X_k = x_k$, denoted by $\triangleleft(X_k = x_k)M$, is defined as follows:

$$\triangleleft(X_k = x_k)\begin{bmatrix} \mu_j \\ \Sigma_{ij} \end{bmatrix}_{1-n} = \begin{bmatrix} \tilde{\mu}_{j.k} \\ \tilde{\Sigma}_{ij.k} \end{bmatrix}_{1-n},$$

where

$$\tilde{\mu}_{j.k} = \begin{cases} \mu_j - (\mu_k + x_k\Sigma_{kk})(\Sigma_{kk})^{-1}\Sigma_{kj}, & j \neq k, \\ -\mu_k(\Sigma_{kk})^{-1}, & j = k, \end{cases}$$

$$\tilde{\Sigma}_{ij.k} = \begin{cases} -(\Sigma_{kk})^{-1}, & i = j = k, \\ -\Sigma_{ik}(\Sigma_{kk})^{-1}, & j = k \neq i, \\ -(\Sigma_{kk})^{-1}\Sigma_{kj}, & i = k \neq j, \\ \Sigma_{ij} - \Sigma_{ik}(\Sigma_{kk})^{-1}\Sigma_{kj}, & \text{otherwise.} \end{cases}$$

Note the forward sweep and the reverse sweep operations defined in [5] are actually the special cases of the above definition when $X_k = 0$. According to multivariate statistics, $\mu_{j.k}$ and $\Sigma_{jj.k}$ $(j \neq k)$ are respectively, the conditional mean and covariance matrix of $X_j$ given $X_k = x_k$. $\Sigma_{jk.k}$ $(j \neq k)$ is the regression coefficient of $X_j$ on $X_k$. Therefore, if $M$ represents a joint distribution, the marginal of $\triangleright(X_k = x_k)M$ to $X_j$ is an extended matrix for the conditional distribution of $X_j$ given $X_k = x_k$.

**Example 3.1.** A Gaussian belief function bears on two deterministic variables $D$ and $U$ and two random variables $X$ and $Y$ as shown by Fig. 1. The relationship among the variables are captured by Eqs. (18) and (19) as follows:

$$X = U + 2Y + \epsilon_X, \tag{18}$$

$$Y = 1 + 2D + \epsilon_Y, \tag{19}$$

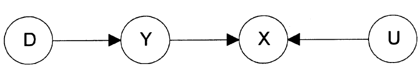


Fig. 1. The probabilistic dependence for $X$, $Y$, $D$, and $U$.

where $\epsilon_X$ and $\epsilon_Y$ are independent residual noises with $\epsilon_X \sim N(0,4)$ and $\epsilon_Y \sim N(0,1)$. For notational simplicity, let us assume that $D = 1$ and $U = 2$. Then by Eqs. (18) and (19), $E(Y) = 1 + 2 + E(\epsilon_Y) = 3$ and $E(X) = 2 + 2E(Y) + E(\epsilon_X) = 8$. By the assumption that $\epsilon_X$ and $\epsilon_Y$ are independent, we have

$$\begin{aligned}
&\mathrm{Var}(Y) = \mathrm{Var}(\epsilon_Y) = 1,\\
&\mathrm{Var}(X) = 4\mathrm{Var}(Y) + \mathrm{Var}(\epsilon_X) = 8,\\
&\mathrm{Cov}(X,Y) = \mathrm{Cov}(U + 2Y + \epsilon_X, Y)\\
&\quad = \mathrm{Cov}(U + 2 + 4D + 2\epsilon_Y + \epsilon_X, 1 + 2D + \epsilon_Y)\\
&\quad = 2\mathrm{Var}(\epsilon_Y, \epsilon_Y) = 2.
\end{aligned}$$

Therefore, the random variables $X$ and $Y$ have extended matrix:

$$\begin{pmatrix} 8 & 8 & 2 \\ 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}}.$$

Applying the forward sweep on $X = 4$ results in

$$\rhd(X=4)\begin{pmatrix} 8 & 8 & 2 \\ 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 1 & -1/8 & 1/4 \\ 2 & 1/4 & 1/2 \end{pmatrix}^{\mathrm{T}}.$$

Therefore, the conditional mean of $Y$ given $X = 4$ is 2 and the conditional variance is $1/2$. Applying the reverse sweep to the above swept matrix yields

$$\lhd(X=4)\begin{pmatrix} 1 & -1/8 & 1/4 \\ 2 & 1/4 & 1/2 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 8 & 8 & 2 \\ 3 & 2 & 1 \end{pmatrix}^{\mathrm{T}}.$$

We note that the reverse sweep nullifies the effect of the forward sweep in Example 3.1. As shown by Lemma 3.1, it is true in general that the forward and reverse sweepings cancel each other.

**Lemma 3.1.** *Suppose $M$ is an extended matrix for set $X$ and $Y$ is a subset of $X$. Then*

$$\lhd(Y=y)[\rhd(Y=y)M] = \rhd(Y=y)[\lhd(Y=y)M] = M. \tag{20}$$

**Proof.** Let $X$ be partitioned into $[X_1, X_2]$ such that $Y = X_2$. Assume

$$M = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}_{1,2},$$

$$\lhd(X_2=y)[\rhd(X_2=y)M] = \begin{pmatrix} \hat{\mu}_{j.2.2} \\ \hat{\Sigma}_{ij.2.2} \end{pmatrix}_{1,2},$$

$$\rhd(X_2 = y)[\lhd(X_2 = y)M] = \begin{pmatrix} \check{\mu}_{j.2.2} \\ \check{\Sigma}_{ij.2.2} \end{pmatrix}_{1,2}.$$

Then we need to show that the following equalities hold for all $i, j = 1, 2$:

$$\hat{\Sigma}_{ij.2.2} = \check{\Sigma}_{ij.2.2} = \Sigma_{ij}, \tag{21}$$

$$\hat{\mu}_{j.2.2} = \check{\mu}_{j.2.2} = \mu_j. \tag{22}$$

Eqs. (21) and (22) can be verified case by case using the definitions of sweepings. For example,

$$\begin{aligned}\hat{\mu}_{1.2.2} = \mu_1 &- (\mu_2 - y)(\Sigma_{22})^{-1}\Sigma_{21} \\ &- \Big\{\mu_2(\Sigma_{22})^{-1} + y[-(\Sigma_{22})^{-1}]\Big\}[-(\Sigma_{22})^{-1}]^{-1}(\Sigma_{22})^{-1}\Sigma_{21} = \mu_1,\end{aligned}$$

$$\begin{aligned}\check{\mu}_{1.2.2} = \mu_1 &- (\mu_2 + y\Sigma_{22})(\Sigma_{22})^{-1}\Sigma_{21} \\ &- [-\mu_2(\Sigma_{22})^{-1} - y][-(\Sigma_{22})^{-1}]^{-1}(-\Sigma_{22})^{-1}\Sigma_{21} = \mu_1.\end{aligned}$$ □

The semantics of $\lhd$ can be understood from Lemma 3.1, which states that $\lhd(Y = y)$ and $\rhd(Y = y)$ nullify each other. As we know, $\rhd(X_2 = y)M$ represents the conditional probability of $X_1$ given $X_2 = y$. Lemma 3.1 implies that $\lhd(X_2 = y)[\rhd(X_2 = y)M]$ recovers the joint probability of $X_1$ and $X_2$. Therefore, applying $\lhd(X_2 = y)$ to $\rhd(X_2 = y)M$ sounds like multiplying the marginal density function of $X_2$ to the conditional density function of $X_1$ given $X_2 = y$. However, it seems difficult to interpret $\rhd(X_2 = y)[\lhd(X_2 = y)M]$ because there are no statistical semantics for the operations $\lhd(X_2 = y)M$ when $M$ represents a joint distribution. However, it is interesting to note that applying $\rhd(X_2 = y)$ to $\lhd(X_2 = y)M$ recovers $M$.

**Lemma 3.2.** *Suppose $M$ is an extended matrix for $X$. $Z$ is an arbitrary set of variables. $Y$ is a subset of both $X$ and $Z$. Then*

$$\rhd(Y = y)M^{\downarrow Z} = [\rhd(Y = y)M]^{\downarrow Z},$$

$$\lhd(Y = y)M^{\downarrow Z} = [\lhd(Y = y)M]^{\downarrow Z}.$$

**Proof.** Let us partition $X$ into $[X_1, X_2, X_3]$ such that $Y = X_1$, $X \cap Z - Y = X_2$, and $X - X \cap Z = X_3$. Note that $\mu_{j.1}$, $\Sigma_{ij.1}$, $\tilde{\mu}_{j.1}$, and $\tilde{\Sigma}_{ij.1}$, $(i, j = 1, 2)$, depend only on $\mu_1$, $\mu_2$, $\Sigma_{11}$, $\Sigma_{12}$, $\Sigma_{21}$, and $\Sigma_{22}$. Lemma 3.2 can then be easily verified by applying the definitions of the forward and reverse sweepings as well as marginalization. □

**Lemma 3.3.** *Suppose $M$ is an extended matrix for $X$ and $Y$ and $Z$ are two disjoint subsets of $X$. Then the forward sweepings on $Y = y$ and on $Z = z$ satisfy the following*:

$$\{\rhd(Y = y)[\rhd(Z = z)M]\}^{\downarrow X-Y\cup Z} = [\rhd(Z = z, Y = y)M]^{\downarrow X-Y\cup Z}, \tag{23}$$

$$\rhd(Y = 0)[\rhd(Z = 0)M] = \rhd(Y = 0, Z = 0)M. \tag{24}$$

**Proof.** Let us partition $X$ into $[X_1, X_2, X_3]$ such that $X_1 = Y$, $X_2 = Z$ and $X_3 = X - Y \cup Z$. Thus, we can write

$$M = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}_{1-3}.$$

We need to prove

$$\{\rhd(X_1 = x_1)[\rhd(X_2 = x_2)M]\}^{\downarrow X_3} = [\rhd(X_1 = x_1, X_2 = x_2)M]^{\downarrow X_3}.$$

In other words, we need to show that $\mu_{3.12} = \mu_{3.1.2}$ and $\Sigma_{33.12} = \Sigma_{33.1.2}$. By definition,

$$\begin{aligned}\mu_{3.12} &= \mu_3 + (x_1 - \mu_1, x_2 - \mu_2)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma_{13} \\ \Sigma_{23} \end{pmatrix} \\ &= \mu_3 + (x_1 - \mu_1)A^{-1}[\Sigma_{13} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{23}] \\ &\quad + (x_2 - \mu_2)B^{-1}[\Sigma_{23} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{13}],\end{aligned}$$

where $A = \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}$ and $B = \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}$. Therefore,

$$\begin{aligned}&A^{-1}[\Sigma_{13} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{23}] \\ &= [(\Sigma_{11})^{-1} + (\Sigma_{11})^{-1}\Sigma_{12}B^{-1}\Sigma_{21}(\Sigma_{11})^{-1}][\Sigma_{13} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{23}] \\ &= (\Sigma_{11})^{-1}\Sigma_{13} + (\Sigma_{11})^{-1}\Sigma_{12}B^{-1}\Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{13} \\ &\quad - (\Sigma_{11})^{-1}\Sigma_{12}[(\Sigma_{22})^{-1} + B^{-1}\Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}(\Sigma_{22})^{-1}]\Sigma_{23} \\ &= (\Sigma_{11})^{-1}\Sigma_{13} - (\Sigma_{11})^{-1}\Sigma_{12}B^{-1}[\Sigma_{23} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{13}].\end{aligned}$$

Therefore,

$$\begin{aligned}\mu_{3.12} &= \mu_3 + (x_1 - \mu_1)(\Sigma_{11})^{-1}\Sigma_{13} \\ &\quad - [\mu_2 + (x_2 - \mu_2)(\Sigma_{11})^{-1}\Sigma_{13} - x_2]B^{-1}[\Sigma_{23} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{13}],\end{aligned}$$

which, by definition, is just $\mu_{3.1.2}$. We can similarly verify $\Sigma_{33.12} = \Sigma_{33.1.2}$ by letting $\mu_3$ be replaced by $\Sigma_{33}$, $x_1 - \mu_1$ by $-\Sigma_{31}$, and $x_2 - \mu_2$ by $-\Sigma_{32}$. Therefore, Eq. (23) is proved.

To prove Eq. (24), we only need to verify $(\mu_{1.12}, \mu_{2.12}) = (\mu_{1.1.2}, \mu_{2.1.2})$. By the definition of the forward sweep, it is easy to see

$$(\mu_{1.12}, \mu_{2.12}) = (\mu_1, \mu_2)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1}$$
$$= ([\mu_1 - \mu_2(\Sigma_{22})^{-1}\Sigma_{21}]A^{-1}, [\mu_2 - \mu_1(\Sigma_{11})^{-1}\Sigma_{12}]B^{-1})$$

and $[\mu_2 - \mu_1(\Sigma_{11})^{-1}\Sigma_{12}]B^{-1} = \mu_{2.1}(\Sigma_{22.1})^{-1} = \mu_{2.1.2}$. Applying transformation $A^{-1} = [(\Sigma_{11})^{-1} + (\Sigma_{11})^{-1}\Sigma_{12}B^{-1}\Sigma_{21}(\Sigma_{11})^{-1}]$, we can also verify that

$$[\mu_1 - \mu_2(\Sigma_{22})^{-1}\Sigma_{21}]A^{-1} = \mu_{1.1} - \mu_{2.1}(\Sigma_{22.1})^{-1}\Sigma_{21.1} = \mu_{1.1.2}.$$

Other equalities can be similarly proved. □

**Lemma 3.4.** *Suppose $M_1$ and $M_2$ are extended matrices, respectively for $[X, Y]$ and $[X, Z]$ with $Y \cap Z = \phi$. Then*

$$\begin{aligned} &\triangleleft(X = x)[\triangleright(X = x)M_1 \oplus \triangleright(X = x)M_2] \\ &\quad = \triangleleft(X = 0)[\triangleright(X = 0)M_1 \oplus \triangleright(X = 0)M_2]. \end{aligned}$$

**Proof.** Let $X_1 = X$, $X_2 = Y$, and $X_3 = Z$. Let

$$M = \triangleright(X = x)M_1 \oplus \triangleright(X = x)M_2,$$
$$\bar{M} = \triangleright(X = 0)M_1 \oplus \triangleright(X = 0)M_2,$$
$$M_1 = \begin{pmatrix} \mu_j^1 \\ \Sigma_{ij}^1 \end{pmatrix}_{1,2}, \quad M_2 = \begin{pmatrix} \mu_j^2 \\ \Sigma_{ij}^2 \end{pmatrix}_{1,3},$$
$$M = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}_{1\text{–}3}, \quad \bar{M} = \begin{pmatrix} \bar{\mu}_j \\ \bar{\Sigma}_{ij} \end{pmatrix}_{1\text{–}3}.$$

Note that $x$ appears only in $\mu_{2.1}$ and $\mu_{3.1}$. Thus, $\mu_{1.1} = \bar{\mu}_{1.1}$ and $\Sigma_{ij.1} = \bar{\Sigma}_{ij.1}$ $(i, j = 1, 2, 3)$. Therefore, we only need to show $\mu_{2.1} = \bar{\mu}_{2.1}$ and $\mu_{3.1} = \bar{\mu}_{3.1}$. By definition, $\Sigma_{12} = (\Sigma_{11}^1)^{-1}\Sigma_{12}^1$ and $\mu_2 = \mu_2^1 - (\mu_1^1 - x)(\Sigma_{11}^1)^{-1}\Sigma_{12}^1$. Therefore,

$$\begin{aligned} \mu_{2.1} &= \mu_2 - (\mu_1 + x\Sigma_{11})(\Sigma_{11})^{-1}\Sigma_{12} \\ &= \mu_2^1 - (\mu_1^1 - x)(\Sigma_{11}^1)^{-1}\Sigma_{12}^1 - (\mu_1 + x\Sigma_{11})(\Sigma_{11})^{-1}(\Sigma_{11}^1)^{-1}\Sigma_{12}^1 \\ &= \mu_2^1 - \mu_1^1(\Sigma_{11}^1)^{-1}\Sigma_{12}^1 - \mu_1(\Sigma_{11})^{-1}(\Sigma_{11}^1)^{-1}\Sigma_{12}^1 \\ &= \mu_2^1 - (\mu_1^1 - 0)(\Sigma_{11}^1)^{-1}\Sigma_{12}^1 - (\mu_1 + 0\Sigma_{11})(\Sigma_{11})^{-1}(\Sigma_{11}^1)^{-1}\Sigma_{12}^1, \end{aligned}$$

which is just $\bar{\mu}_{2.1}$. Similarly we can verify that $\mu_{3.1} = \bar{\mu}_{3.1}$. □

We now use the sweeping operators to re-express Dempster's rule for combining GBFs obtained in Liu [14]. Given two GBFs $\mathrm{Bel}_1 = (\mathbf{C}_1, \mathbf{B}_1, \pi_1, E_1)$ and $\mathrm{Bel}_2 = (\mathbf{C}_2, \mathbf{B}_2, \pi_2, E_2)$, without loss of generality, assume we can choose a convenient basis $D_1$, $D_2$, $D_3$, $U$, $V$, $X_1$, $X_2$, $X_3, \ldots$ such that $\mathbf{C}_1$ is spanned by $\{D_1, D_2, U\}$, $\mathbf{C}_2$ by $\{D_1, D_3, V\}$, $\mathbf{B}_1$ by $\{D_1, D_2, U, V, X_1, X_2\}$, and $\mathbf{B}_2$ by

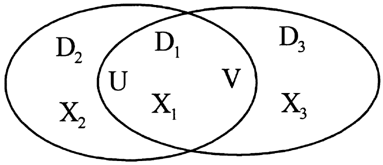


Fig. 2. The hypergraph for $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$.

$\{D_1, D_3, V, U, X_1, X_3\}$. The hypergraph representing $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$ is shown in Fig. 2. As we can see, $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$ share four sets of common variables: $D_1$, $X_1$, $U$, and $V$. $D_1$ is deterministic and $X_1$ is uncertain to both $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$. $U$ is deterministic in $\mathrm{Bel}_1$ but uncertain in $\mathrm{Bel}_2$. $V$ is otherwise. In addition, $\mathrm{Bel}_1$ is certain about $D_2$ and uncertain about $X_2$. However, $\mathrm{Bel}_2$ has no opinions about either $D_2$ or $X_2$. Similarly, $\mathrm{Bel}_2$ is certain about $D_3$ and uncertain about $X_3$ but $\mathrm{Bel}_1$ has no opinions about either of them.

**Theorem 3.1.** *Given any two GBFs* $\mathrm{Bel}_1 = (\mathbf{C}_1, \mathbf{B}_1, \pi_1, E_1)$: $\mathbf{C}_1$ *is spanned by* $\{D_1, D_2, U\}$ *with* $D_1 = d_1$, $D_2 = d_2$, $U = u$, *and* $\mathbf{B}_1$ *by* $\{D_1, D_2, U, V, X_1, X_2\}$ *and* $\mathrm{Bel}_2 = (\mathbf{C}_2, \mathbf{B}_2, \pi_2, E_2)$: $\mathbf{C}_2$ *is spanned by* $\{D_1, D_3, V\}$ *with* $D_1 = d_1$, $D_3 = d_3$, $V = v$, $\mathbf{B}_2$ *by* $\{D_1, D_3, V, U, X_1, X_3\}$, *their combination* $\mathrm{Bel}_1 \otimes \mathrm{Bel}_2$ *is* $\mathrm{Bel} = (\mathbf{C}, \mathbf{B}, \pi, E)$: $\mathbf{C} = \mathbf{C}_1 \oplus \mathbf{C}_2$ *with* $D_1 = d_1$, $D_2 = d_2$, $D_3 = d_3$, $U = u, V = v$, $\mathbf{B} = \mathbf{B}_1 \oplus \mathbf{B}_2$. *Let* $M_1$, $M_2$, *and* $M$ *as follows be respectively*, *the extended matrices of* $\mathrm{Bel}_1$, $\mathrm{Bel}_2$ *and* Bel:

$$M_1 = \begin{pmatrix} \mu_j^1 \\ \Sigma_{ij}^1 \end{pmatrix}_{0,1,2}, \quad M_2 = \begin{pmatrix} \mu_j^2 \\ \Sigma_{ij}^2 \end{pmatrix}_{0,1,3}, \quad M = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}_{1\text{–}3}.$$

*Then*

$$M = \lhd(X_1 = 0)\Big\{ \rhd (X_1 = 0)[\rhd(V = v)M_1]^{\downarrow[X_1, X_2]} \oplus \rhd(X_1 = 0) \times [\rhd(U = u)M_2]^{\downarrow[X_1, X_3]}\Big\}.$$

**Proof.** First we note that $D_1 = d_1$ in both $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$. Since $D_1$ is deterministic, its value must be the same in both $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$ otherwise they have not common ground for agreements to be combinable. According to Liu [14], $\mathbf{C} = \mathbf{C}_1 \oplus \mathbf{C}_2$ and $\mathbf{B} = \mathbf{B}_1 \oplus \mathbf{B}_2$. Therefore, what we need to show is to verify that $M$ is the extended matrix for $\mathrm{Bel}_1 \otimes \mathrm{Bel}_2$.

According to Liu [14], $\mathrm{Bel}_1 \otimes \mathrm{Bel}_2$ has extended matrix:

$$M = \begin{pmatrix} a_1 & a_2 + a_1(b_2)^{\mathrm{T}} & a_3 + a_1(b_3)^{\mathrm{T}} \\ \sigma_1 & \sigma_1(b_2)^{\mathrm{T}} & \sigma_1(b_3)^{\mathrm{T}} \\ b_2\sigma_1 & \sigma_2 + b_2\sigma_1(b_2)^{\mathrm{T}} & b_2\sigma_1(b_3)^{\mathrm{T}} \\ b_3\sigma_1 & b_3\sigma_1(b_2)^{\mathrm{T}} & \sigma_3 + b_3\sigma_1(b_3)^{\mathrm{T}} \end{pmatrix}, \tag{25}$$

where

$$\sigma_1 = [(\Sigma^1)^{-1} + (\Sigma^2)^{-1}]^{-1}, \tag{26}$$

$$\sigma_2 = \Sigma^1_{22} - (\Sigma^1_{20}, \Sigma^1_{21}) \begin{pmatrix} \Sigma^1_{00} & \Sigma^1_{01} \\ \Sigma^1_{10} & \Sigma^1_{11} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma^1_{02} \\ \Sigma^1_{12} \end{pmatrix}, \tag{27}$$

$$\sigma_3 = \Sigma^2_{33} - (\Sigma^2_{30}, \Sigma^2_{31}) \begin{pmatrix} \Sigma^2_{00} & \Sigma^2_{01} \\ \Sigma^2_{10} & \Sigma^2_{11} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma^2_{03} \\ \Sigma^2_{13} \end{pmatrix}, \tag{28}$$

$$a_1 = [\mu^1(\Sigma^1)^{-1} + \mu^2(\Sigma^2)^{-1}][(\Sigma^1)^{-1} + (\Sigma^2)^{-1}]^{-1}. \tag{29}$$

$\Sigma^1$, $\Sigma^2$, $\mu^1$, and $\mu^2$ in Eqs. (26) and (29) are determined by the following:

$$\Sigma^i = \Sigma^i_{11} - \Sigma^i_{10}(\Sigma^i_{00})^{-1}\Sigma^i_{01} \quad (i = 1, 2), \tag{30}$$

$$\mu^1 = \mu^1_1 + (v - \mu^1_0)(\Sigma^1_{00})^{-1}\Sigma^1_{01}, \tag{31}$$

$$\mu^2 = \mu^2_1 + (u - \mu^2_0)(\Sigma^2_{00})^{-1}\Sigma^2_{01}. \tag{32}$$

$a_i$ and $b_i$ $(i = 2, 3)$ in Eq. (25) are implicitly determined by matching the coefficients of $x_1$ on the both sides of the following equations:

$$a_2 + x_1(b_2)^{\mathrm{T}} = \mu^1_2 - (v - \mu^1_0, x_1 - \mu^1_1) \begin{pmatrix} \Sigma^1_{00} & \Sigma^1_{01} \\ \Sigma^1_{10} & \Sigma^1_{11} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma^1_{02} \\ \Sigma^1_{12} \end{pmatrix}, \tag{33}$$

$$a_3 + x_1(b_3)^{\mathrm{T}} = \mu^2_3 + (u - \mu^2_0, x_1 - \mu^2_1) \begin{pmatrix} \Sigma^2_{00} & \Sigma^2_{01} \\ \Sigma^2_{10} & \Sigma^2_{11} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma^2_{03} \\ \Sigma^2_{13} \end{pmatrix}. \tag{34}$$

Note that the superscripts 1 and 2, respectively correspond to $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$. The subscript 0 in $\mathrm{Bel}_1$ corresponds to the variable $V$, which is deterministic in $\mathrm{Bel}_2$. In contrast, the subscript 0 in $\mathrm{Bel}_2$ corresponds to the variable $U$, which is deterministic in $\mathrm{Bel}_1$.

By the definition of the reverse sweep, we can verify the following:

$$\begin{aligned} M &= \triangleleft(X_1 = x_1) \begin{pmatrix} \mu^1(\Sigma^1)^{-1} + \mu^2(\Sigma^2)^{-1} & a_2 + x_1(b_2)^{\mathrm{T}} & a_3 + x_1(b_3)^{\mathrm{T}} \\ -(\Sigma^1)^{-1} - (\Sigma^2)^{-1} & (b_2)^{\mathrm{T}} & (b_3)^{\mathrm{T}} \\ b_2 & \sigma_2 & 0 \\ b_3 & 0 & \sigma_3 \end{pmatrix} \\ &= \triangleleft(X_1 = x_1) \left[ \begin{pmatrix} \mu^1(\Sigma^1)^{-1} & a_2 + x_1(b_2)^{\mathrm{T}} \\ -(\Sigma^1)^{-1} & (b_2)^{\mathrm{T}} \\ b_2 & \sigma_2 \end{pmatrix} \oplus \begin{pmatrix} \mu^2(\Sigma^2)^{-1} & a_3 + x_1(b_3)^{\mathrm{T}} \\ -(\Sigma^2)^{-1} & (b_3)^{\mathrm{T}} \\ b_3 & \sigma_3 \end{pmatrix} \right]. \end{aligned}$$

Then, plugging Eqs. (25)–(34), we can verify that

$$\triangleright (X_1 = x_1)[\triangleright(V = v)M_1]^{\downarrow[X_1,X_2]} = \begin{pmatrix} \mu^1(\Sigma^1)^{-1} & a_2 + x_1(b_2)^{\mathrm{T}} \\ -(\Sigma^1)^{-1} & (b_2)^{\mathrm{T}} \\ b_2 & \sigma_2 \end{pmatrix},$$

$$\triangleright (X_1 = x_1)[(U = u)M_2]^{\downarrow[X_1,X_3]} = \begin{pmatrix} \mu^2(\Sigma^2)^{-1} & a_3 + x_1(b_3)^{\mathrm{T}} \\ -(\Sigma^2)^{-1} & (b_3)^{\mathrm{T}} \\ b_3 & \sigma_3 \end{pmatrix}.$$

Therefore, by Lemma 3.4, Theorem 3.1 is proved. □

Theorem 3.1 simplifies a complex process of combining GBFs into one well organized formula. Since sweepings can be easily done using spreadsheets, the theorem implies that combining GBFs can be as simple as computing basic statistics. As we will see in Section 4, Theorem 3.1 also plays an important role in proving the feasibility of the join-tree computation scheme for GBFs.

**Example 3.2.** Let $\mathrm{Bel}_1$ be the GBF in Example 3.1 and $M_1$ be its extended matrix. $\mathrm{Bel}_2$ be another GBF bearing on random variables $U$, $X$, and $Z$, represented by the following statistical models:

$$Z = 1.5U + 0.8 + \epsilon_Z, \tag{35}$$

$$X = Z + 5 + \epsilon_X, \tag{36}$$

where $U \sim N(1.9, 0.04)$, $\epsilon_Z \sim N(0, 2)$, and $\epsilon_X \sim N(0, 6)$ are independent. The belief network representation of the above model is shown in Fig. 3. Then we can show that the variables $U$, $X$ and $Z$ has an extended matrix as follows:

$$M_2 = \begin{pmatrix} 1.90 & 0.04 & 0.06 & 0.06 \\ 8.65 & 0.06 & 8.09 & 2.09 \\ 3.65 & 0.06 & 2.09 & 2.09 \end{pmatrix}^{\mathrm{T}}.$$

$U$ is random in $\mathrm{Bel}_2$ but deterministic in $\mathrm{Bel}_1$. On the other hand, $\mathrm{Bel}_1$ has no random variables that are deterministic in $\mathrm{Bel}_2$. Thus, according to Theorem 3.1, we only need to apply the forward sweep to $M_2$ on $U = 2$:

$$[\triangleright(U = 2)M_2]^{\downarrow\{X,Z\}} = \begin{pmatrix} 8.8 & 8 & 2 \\ 3.8 & 2 & 2 \end{pmatrix}^{\mathrm{T}}.$$

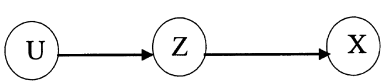


Fig. 3. The Gaussian belief network for Eqs. (35) and (36).

We then apply $\triangleright(X=0)$ to $M_1$ and $[\triangleright(U=2)M_2]^{\downarrow\{X,Z\}}$:

$$\triangleright(X=0)M_1 = \begin{pmatrix} 1 & -1/8 & 1/4 \\ 1 & 1/4 & 1/2 \end{pmatrix}^{\mathrm{T}},$$

$$\triangleright(X=0)[\triangleright(U=2)M_2]^{\downarrow\{X,Z\}} = \begin{pmatrix} 1.1 & -1/8 & 1/4 \\ 1.6 & 1/4 & 3/2 \end{pmatrix}^{\mathrm{T}}.$$

We add the above two matrices using the direct sum:

$$\triangleright(X=0)M_1 \oplus \triangleright(X=0)[\triangleright(U=2)M_2]^{\downarrow\{X,Z\}} = \begin{pmatrix} 2.1 & -1/4 & 1/4 & 1/4 \\ 1.0 & 1/4 & 1/2 & 0 \\ 1.6 & 1/4 & 0 & 3/2 \end{pmatrix}.$$

Finally, we apply $\triangleleft(X=0)$ to the resulting matrix and obtain the following:

$$\begin{pmatrix} 8.4 & 4 & 1 & 1 \\ 3.1 & 1 & 0.75 & 0.25 \\ 3.7 & 1 & 0.25 & 1.75 \end{pmatrix}^{\mathrm{T}},$$

which is the extended matrix for $\mathrm{Bel}_1 \otimes \mathrm{Bel}_2$ according to Theorem 3.1.

Given $D=1$ and $U=2$, according to Eqs. (18) and (19), the predicted value of $X$ is 8 with variance 8. According to Eqs. (35) and (36), the predicted value of $X$ is 8.8 with variance 8. In the combined model, the predicted value of $X$ is 8.4 with variance 4. According to Eq. (29), the mean of $X$ in the combined model is the inverse variance-weighted average of the means of $X$ in the component models. For any component model, the smaller the variance, the larger the weight associated with the mean. This is reasonable. If a component model is subjectively assessed, a small variance means the high certainty of a subjective belief. If a component model is statistically estimated, a small variance corresponds to the large sample size. In either case, the estimation corresponding to a small variance should be given a large weight. In this perspective, Eq. (29) have some flavor of meta-analysis, where sample sizes are often taken as a principal weighting factor. In Example 3.2, we also note that the variance of $X$ in Bel is smaller than that in both $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$. However, the variance of $Y$ in Bel is larger than that in $\mathrm{Bel}_1$ and the variance of $Z$ in Bel is larger than that in $\mathrm{Bel}_2$. It turns out that this is generally true. As Liu [12] shows, the opinions on a common variable can strengthen each other in the sense that the aggregate opinion on the variable tends to be more focused than individual ones. In contrast, the opinion on a variable expressed by a single GBF tends to be eroded by another GBF that has no opinion on the variable in the sense that the combination makes the variance of the variable increase.

## 4. Propagation in join-trees

Suppose we have many models to be combined but we are only interested in making inference for a subset of variables involved. An obvious approach to the problem is to combine all the GBFs into one GBF and then marginalize it down to the variables of interest. However, it is equally obvious to see that such an approach may be very inefficient when the number of the variables involved is very large. An alternative approach is to apply a join-tree algorithm and combine GBFs locally. Join-tree algorithms have been very successful for many diverse problem domains including uncertainty reasoning in expert systems [10,26], the management of relational databases [15], and the solution of influence diagrams [7,25]. The basic idea of the approach is to arrange all the variables into a tree-structured graph, called *a join-tree*, and propagate knowledge by sending and absorbing messages step-by-step in the tree. Each step involves sending a message from a node to a neighbor and thus involves only a small number of variables that are near each other in the join-tree.

The join-tree approach is shown to be applicable to finite and to condensable belief functions [9,24,28]. There is some work yet to be done, however, in justifying the approach in the case of GBFs that are neither finite nor condensable. According to Shenoy and Shafer [28], the justification amounts to proving that combination and marginalization of GBFs follow the Shafer–Shenoy axioms, which are the conditions under which exact local computation is possible.

**Theorem 4.1.** *The following three axioms of Shenoy and Shafer* [28] *hold for GBFs*:

**Axiom 1.** Combination of GBFs is commutative and associative.

**Axiom 2.** Let $\mathrm{Bel} = (\mathbf{C}, \mathbf{B}, \pi, E)$ and $\mathbf{K} \subset \mathbf{M}$. Then

$$(\mathrm{Bel}^{\downarrow \mathbf{M}})^{\downarrow \mathbf{K}} = \mathrm{Bel}^{\downarrow \mathbf{K}}.$$

**Axiom 3.** Let $\mathrm{Bel}_1 = (\mathbf{C}_1, \mathbf{B}_1, \pi_1, E_1)$ and $\mathrm{Bel}_2 = (\mathbf{C}_2, \mathbf{B}_2, \pi_2, E_2)$. Then

$$(\mathrm{Bel}_1 \otimes \mathrm{Bel}_2)^{\downarrow \mathbf{B}_1} = \mathrm{Bel}_1 \otimes (\mathrm{Bel}_2)^{\downarrow \mathbf{B}_1}.$$

**Proof.** Since intersection and addition are commutative and associative, Eq. (1) implies that Axiom 1 holds. Axiom 2 can be also easily verified by the marginalization rule defined in Eq. (2). Thus we only need to verify Axiom 3, the distributivity of marginalization over combination. Let $\mathrm{Bel} = \mathrm{Bel}_1 \otimes \mathrm{Bel}_2 = (\mathbf{C}, \mathbf{B}, \pi, E)$, where $\mathbf{C} = \mathbf{C}_1 \cap \mathbf{C}_2$ and $\mathbf{B} = \mathbf{B}_1 \cap \mathbf{B}_2$. Therefore, the certainty and the belief spaces for $(\mathrm{Bel}_1 \otimes \mathrm{Bel}_2)^{\downarrow \mathbf{B}_1}$ are respectively, $\mathbf{C}_1 \cap \mathbf{C}_2 \cap (\mathbf{B}_1) = \mathbf{C}_1 \cap (\mathbf{C}_2 \cap \mathbf{B}_1)$ and $\mathbf{B}_1 \cap \mathbf{B}_2 \cap (\mathbf{B}_1) = \mathbf{B}_1 \cap (\mathbf{B}_2 \cap \mathbf{B}_1)$, which can be easily seen to be the certainty and the belief spaces for $\mathrm{Bel}_1 \otimes (\mathrm{Bel}_2)^{\downarrow \mathbf{B}_1}$. Therefore, to show Axiom 3, we only need to prove that both $(\mathrm{Bel}_1 \otimes \mathrm{Bel}_2)^{\downarrow \mathbf{B}_1}$ and $\mathrm{Bel}_1 \otimes (\mathrm{Bel}_2)^{\downarrow \mathbf{B}_1}$ have the same extended matrix. Suppose the representation matrices of Bel, $\mathrm{Bel}_1$, and $\mathrm{Bel}_2$ are respectively, $M$, $M_1$, and $M_2$. Without loss of generality, let us assume $\mathbf{B}_1$ is spanned by the variables that can be partitioned into $[U, V, X_1, X_2]$, where $U$ is a set of deterministic variables with $U = u$, and $\mathbf{B}_2$ is spanned by the variable that can be partitioned into $[U, V, X_1, X_3]$, where $V$ is a set of deterministic variables with $V = v$. Then, according to Theorem 3.1,

$$
\begin{aligned}
M = {} & \triangleleft(X_1 = 0) \\
& \left\{ \triangleright (X_1 = 0)[\triangleright(V = v)M_1]^{\downarrow [X_1, X_2]} \oplus \triangleright(X_1 = 0)[\triangleright(U = u)M_2]^{\downarrow [X_1, X_3]} \right\}.
\end{aligned}
$$

According to Lemma 3.2 and Eq. (17), we can verify the following:

$$
\begin{aligned}
M^{\downarrow \mathbf{B}_1} &= M^{\downarrow [U,V,X_1,X_2]} \\
&= \triangleleft (X_1 = 0) \Big\{ \triangleright (X_1 = 0) \Big\{ [\triangleright(V = v)M_1]^{\downarrow [X_1, X_2]} \Big\}^{\downarrow [U,V,X_1,X_2]} \\
&\qquad \oplus \triangleright(X_1 = 0) \Big\{ [\triangleright(U = u)M_2]^{\downarrow [X_1, X_3]} \Big\}^{\downarrow [U,V,X_1,X_2]} \Big\}.
\end{aligned}
$$

Then by Lemma 3.2 and Eq. (16), we have

$$
\begin{aligned}
M^{\downarrow \mathbf{B}_1} &= \triangleleft (X_1 = 0) \Big\{ \triangleright (X_1 = 0)[\triangleright(V = v)M_1]^{\downarrow [X_1, X_2] \cap [U,V,X_1,X_2]} \\
&\qquad\qquad \oplus \triangleright(X_1 = 0)[\triangleright(U = u)M_2]^{\downarrow [X_1, X_3] \cap [U,V,X_1,X_2]} \Big\} \\
&= \triangleleft (X_1 = 0) \Big\{ \triangleright (X_1 = 0)[\triangleright(V = v)M_1]^{\downarrow [X_1, X_2]} \\
&\qquad\qquad \oplus \triangleright(X_1 = 0) \Big\{ [\triangleright(U = u)M_2]^{\downarrow [U,V,X_1,X_2]} \Big\}^{\downarrow X_1} \Big\}.
\end{aligned}
$$

Finally, we apply Lemma 3.2 again to the second part of the above direct sum and obtain:

$$M^{\downarrow \mathbf{B}_1} = \lhd (X_1 = 0)\Big\{ \rhd (X_1 = 0)[\rhd(V = v)M_1]^{\downarrow[X_1,X_2]} \oplus \rhd (X_1 = 0)[\rhd(U = u)(M_2)^{\downarrow[U,V,X_1,X_2]}]^{\downarrow X_1}\Big\}.$$

According to the notion of marginalization,

$$(\mathrm{Bel}_2)^{\downarrow \mathbf{B}_1} = (\mathbf{C}_2 \cap \mathbf{B}_1, \mathbf{B}_2 \cap \mathbf{B}_1, \pi|_{\mathbf{B}_2\cap\mathbf{B}_1}, E|_{\mathbf{B}_2\cap\mathbf{B}_1}).$$

$\mathbf{B}_2 \cap \mathbf{B}_1$ is spanned by $U$, $V$, and $X_1$, where $V$ is deterministic with $V = v$. $(M_2)^{\downarrow[U,V,X_1,X_2]}$ is the extended matrix for $[U, X_1]$. Therefore, $M^{\downarrow[U,V,X_1,X_2]}$ is the same as the extended matrix for $\mathrm{Bel}_1 \otimes (\mathrm{Bel}_2)^{\downarrow B_1}$. Therefore, Axiom 3 is proved. □

Axiom 1 implies that the order of combination does not matter. Thus, if we have many GBFs to be combined, we can write the combination as $\otimes\{\mathrm{Bel}_i|i\}$. Axiom 2 implies the transitivity of marginalization. Axiom 3 implies that marginalization is distributive over combination. These three axioms jointly imply that combination and marginalization of GBFs can be executed locally in a join-tree. In the following, we adapt the Shafer–Shenoy architecture of local computation to show how this can be done in details.

A join-tree is a tree-structured graph, where each node is a subset of variables, each pair of neighbors has non-empty intersection, and the intersection of two distinct nodes is contained in every node on the path connecting the two distinct nodes. For example, Fig. 4 shows a join-tree for four GBFs. To obtain a join-tree, we first draw a hypergraph, where each set of variables on which a GBF bears in represented by a hyperedge. A hypergraph is then transformed into an undirected graph, called *a bi-section graph* in [9], by connecting every pair of variables in each hyperedge. For example, Fig. 6 shows the bisection graph for the hypergraph in Fig. 5. After this step, a successive procedure, called *filling-in*, is taken to triangulate the undirected graph by adding chords in any chordless cycles of length 4 or more. For example, we can triangulate the network in Fig. 6 by connecting $X_1$ and $X_3$ into Fig. 7. Tarjan and Yannakakis [29] proposes a very fast algorithm, called *maximum cardinality search*, for testing triangulatedness. It assigns number 1 to an arbitrary node, numbers the nodes consecutively by choosing as the next to number a node with a maximum

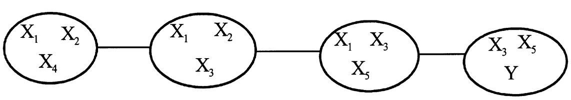


Fig. 4. A join-tree for four GBFs.

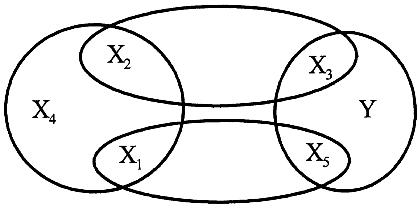


Fig. 5. A hypergraph.

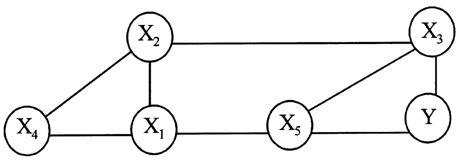


Fig. 6. A bisection graph.

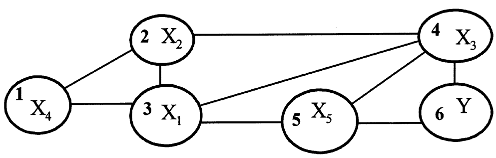


Fig. 7. The triangulation of Fig. 6.

number of previously numbered neighbors, and break ties arbitrarily. If a graph is triangulated, then the numbering will be perfect in the sense that $\mathrm{bd}(i) \cap \{1, 2, \ldots, i-1\}$ is complete for any $i$, where $\mathrm{bd}(i)$ is the set of all the neighbors of $i$. For example, we can check that the numbering attached to Fig. 7 is perfect. In contrast, applying the maximum cardinality search to Fig. 6, we could also obtain the same numbering on Fig. 7. However, in this case, $\mathrm{bd}(5) \cap \{1, 2, 3, 4\} = \{2, 3, 4\}$, which is not complete in Fig. 6.

In an undirected graph, a set of nodes is called a *clique* if it is maximally complete. For example, in Fig. 7, $\{3, 4, 5\}$ is a clique while $\{3, 4\}$ is not. The other three cliques are $\{1, 2, 3\}$, $\{2, 3, 4\}$, and $\{4, 5, 6\}$. Given a numbering produced in triangulation, we can order the cliques by the maximal number assigned in each clique. Let $C_1$, $C_2, \ldots$ denote a list of the ordered cliques. Then, as shown by Tarjan and Yannakakis [29], the cliques satisfy *the running* *intersection property*: For any $i$, $C_i \cap (C_1 \cup C_2 \cup \cdots \cup C_{i-1})$ is contained in a clique among $C_1, C_2, \ldots,$ and $C_{i-1}$. Suppose there exists a $k \leqslant i - 1$ such that $C_i \cap (C_1 \cup C_2 \cup \cdots \cup C_{i-1}) \subseteq C_k$. Then, in assembling a join-tree, we attach $C_i$ to $C_k$. The resulting tree is then a join-tree with nodes $C_1$, $C_2$, .... For example, we can order the four cliques in Fig. 7 as follows:

$$C_1 = \{1,2,3\}, \quad C_2 = \{2,3,4\}, \quad C_3 = \{3,4,5\}, \quad C_4 = \{4,5,6\}.$$

Since $C_2 \cap C_1 \subseteq C_1$, $C_3 \cap (C_1 \cup C_2) \subseteq C_2$, and $C_4 \cap (C_1 \cup C_2 \cup C_3) \subseteq C_3$, we attach $C_2$ to $C_1$, $C_3$ to $C_2$, and $C_4$ to $C_3$. The resulting join-tree is shown in Fig. 4. Note that join-tree representation may not be unique. Different join-trees have different computational advantages and some are better than others. However, finding the best join-tree is NP-complete [1]. There are a number of heuristic methods available for finding reasonably good join-trees [29,9,16,31].

Note that each hyperedge in a hypergraph is a subset of a clique and each hyperedge represents one GBF. Therefore, the GBF associated with a clique is the combination of all the GBFs that bear on the hyperedges of the clique. The join-tree along with associated GBFs is then an equivalent representation of the original hypergraph. However, a join-tree provides a computational structure for more efficient reasoning on the GBFs.

Join-tree algorithms were independently developed in [10] for Bayesian expert systems, in [9,24] for belief function expert systems, in [23] for both Bayesian and belief function systems, and successfully improved in [8]. Here we adapt the Shafer–Shenoy algorithm [22] for computing GBFs. The algorithm regards a marginal GBF as a message and repeatedly apply the following rules to propagate messages:

*Rule* 1. Each node waits to send its message to a given neighbor until it has received messages from all of its other neighbors.

*Rule* 2. When a node is ready to send its message to a particular neighbor, it computes the message by collecting all its messages from other neighbors, combining them with its own GBF, and marginalizing the combined GBF to its intersection with the neighbor to whom it is sending.

Let $C_i$, $C_j$, $C_k$ be clique nodes in a join-tree. Let $m_{i\to j}$ denote the Shafer–Shenoy message to $C_j$ from neighbor $C_i$. Let $\mathrm{Bel}_j$ denote the belief function for clique $C_j$. Let $\mathrm{bd}(C_j)$ denote the set of all neighbors of $C_j$. Then Rule 2 says that the message from $C_j$ to neighbor $C_k$ is given by

$$m_{j\to k} = \left\{\mathrm{Bel}_j \otimes [m_{i\to j} | C_i \in \mathrm{bd}(C_j) - C_k]\right\}^{\downarrow C_j \cap C_k}. \tag{37}$$

Because of Rule 1, the computation must begin with the leaves of a join-tree. A message from a leave is simply the marginal of its own GBF according to Eq. (37). In Fig. 8, for example, the leaves are $C_3$, $C_4$, and $C_5$. Any of these

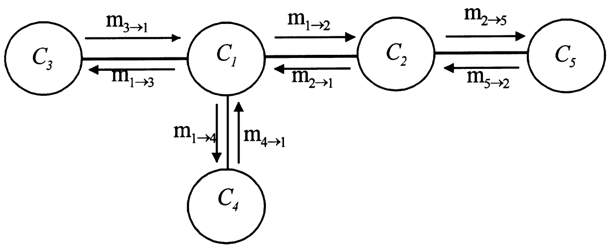


Fig. 8. The Shafer–Shenoy architecture for propagation.

leaves can begin to send their marginals to their unique neighbors. A non-leave node must wait to send its message until it hears from all its other neighbors. In Fig. 8, for example, node $C_1$ sends its message to node $C_2$ only after it receives from $C_3$ and $C_4$ and to $C_3$ only after receiving messages from $C_4$ and $C_2$. Similarly, $C_2$ sends its message to $C_1$ after receiving the message from $C_5$ and to $C_5$ after receiving the message from $C_1$.

The GBF stored in each node does not change during propagation. The message sent from a node to a neighbor is first registered on their communication channel until it is collected in accordance with Rule 2. After all the messages are sent, we have an architecture like Fig. 8, where each node stores its original GBF and all the two-way communication channels are filled with messages. Let Bel denote the combined GBF of all the component GBFs in a join-tree. Then, according to [28],

$$\mathrm{Bel}^{\downarrow C_j} = \mathrm{Bel}_j \otimes \left\{ m_{i\to j} | C_i \in \mathrm{bd}(C_j) \right\}. \tag{38}$$

In words, to obtain the marginal of Bel over the variables at node $C_j$, we collect all the messages to $C_j$ and combine them with $\mathrm{Bel}_j$. Therefore, instead of combining all the GBFs and then marginalizing the combined GBF to $C_j$, Eq. (38) serves the same purpose much more efficiently through local computation. Also, after all the messages are sent, Eq. (38) can be used to compute the marginal GBFs for all the join-tree nodes.

**Example 4.1.** We consider the join-tree computation of five GBFs: $\mathrm{Bel}_1$ and $\mathrm{Bel}_2$ are defined in Example 3.2, $\mathrm{Bel}_3$ is for deterministic variables $D$ and $U$ with observations $D = 1$ and $U = 2$, $\mathrm{Bel}_4$ is for Gaussian variable $X$ with mean 8.5 and variance 2, $\mathrm{Bel}_5$ is for Gaussian variable $Z$ with mean 4 and variance 1. The join-tree representation for these GBFs is shown in Fig. 8, where node $C_i$ has $\mathrm{Bel}_i$, $i = 1, 2, \ldots, 5$. Note that node $C_1$ consists of variables $D$, $U$, $X$, and $Y$ and node $C_2$, variables $U$, $X$, $Z$. By Eq. (37), it is easy to see that

$$m_{3\to 1} = (\mathrm{Bel}_3)^{\downarrow\{D,U\}} = \mathrm{Bel}_3, m_{4\to 1} = (\mathrm{Bel}_4)^{\downarrow X} = \mathrm{Bel}_4$$

and $m_{5\to 2} = (\mathrm{Bel}_5)^{\downarrow Z} = \mathrm{Bel}_5$. Then, by Eq. (37) and Theorem 3.1, we can compute other messages:

$$m_{1\to 2} = \{\mathrm{Bel}_1 \otimes m_{3\to 1} \otimes m_{4\to 1}\}^{\downarrow C_1\cap C_2} = \{\mathrm{Bel}_1 \otimes \mathrm{Bel}_3 \otimes \mathrm{Bel}_4\}^{\downarrow\{U,X\}}.$$

According to Theorem 3.1, $\mathrm{Bel}_1 \otimes \mathrm{Bel}_3 \otimes \mathrm{Bel}_4$ is GBF with $U = 2$ and $X, Y$ have extended matrix

$$\begin{pmatrix} 8.4 & 1.6 & 0.4 \\ 3.1 & 0.4 & 0.6 \end{pmatrix}^{\mathrm{T}}.$$

Thus, $m_{1\to 2}$ is a GBF with $U = 2$ and $X$ has mean 8.4 and variance 1.6. By Theorem 3.1, $\mathrm{Bel}_2 \otimes m_{1\to 2}$ is a GBF with $U = 2$ and $X$ and $Z$ have extended matrix

$$\begin{pmatrix} 8.47 & 4/3 & 1/3 \\ 3.72 & 1/3 & 1.58 \end{pmatrix}^{\mathrm{T}}.$$

Thus, $m_{2\to 5} = \{\mathrm{Bel}_2 \otimes m_{1\to 2}\}^{\downarrow Z}$ is GBF for $Z$ with mean 3.72 and variance 1.58. $m_{2\to 1} = \{\mathrm{Bel}_2 \otimes \mathrm{Bel}_5\}^{\downarrow\{D,X\}}$ has extended matrix

$$\begin{pmatrix} 1.91 & 0.039 & 0.019 \\ 8.89 & 0.019 & 6.676 \end{pmatrix}^{\mathrm{T}}.$$

Since in $\mathrm{Bel}_1$, $D = 1$ and $U = 2$ are certain, $m_{1\to 3} = \{\mathrm{Bel}_1 \otimes m_{2\to 1} \otimes m_{4\to 1}\}^{\downarrow\{D,U\}}$ is GBF with $D = 1$ and $U = 2$. Finally, $m_{1\to 4} = \{\mathrm{Bel}_1 \otimes m_{2\to 1} \otimes m_{3\to 1}\}^{\downarrow X} = \{\mathrm{Bel}_1 \otimes m_{2\to 1}\}^{\downarrow X}$ is GBF for $X$ with mean 8.07 and variance 3.64.

After the propagation, we fill all the communication channels with messages. The GBFs stored in all the nodes do not change. Let $\mathrm{Bel} = \otimes\{\mathrm{Bel}_i \mid i = 1, 2, \ldots, 5\}$. According to Eq. (38), we can compute the marginal of Bel for each node in Fig. 8. For example, $\mathrm{Bel}^{\downarrow X} = \mathrm{Bel}_4 \otimes m_{1\to 4}$ is GBF for $X$ with mean 8.33 and variance 1.29, $\mathrm{Bel}^{\downarrow Z} = \mathrm{Bel}_5 \otimes m_{2\to 5}$ is GBF for $Z$ with mean 3.89 and variance 0.61, and $\mathrm{Bel}^{\downarrow\{D,U,X,Y\}} = \mathrm{Bel}_1 \otimes m_{3\to 1} \otimes m_{4\to 1} \otimes m_{2\to 1}$ is GBF for $\{D, U, X, Y\}$ with $D = 1$, $U = 2$, and $X$ and $Y$ has extended matrix

$$\begin{pmatrix} 8.33 & 1.29 & 0.32 \\ 3.10 & 0.32 & 0.58 \end{pmatrix}^{\mathrm{T}}.$$

Usually, some variables, like dependent variables in a regression model, are of particular interest. The join-tree technique for GBFs allows us to obtain a combined prediction on the variables in accordance with all the models available. In Example 4.1, for instance, the five models have made a combined prediction that the average value of $X$ is 8.33 given $D = 1$ and $U = 2$. The technique also allows us to efficiently absorb evidence and observe its effect throughout the network. For example, suppose we observe that $Z = 3$ in Fig. 8.

We can turn $\mathrm{Bel}_5$ into the GBF for deterministic variable $Z$ with $Z = 3$, propagate the observation through the join-tree as above, and observe the conditional changes of other variables. The join-tree technique can be also used for hypothesizing and planning. Suppose several variables are observable. We can determine which one has more sensitive effect to the variables of interest. By doing so, we can determine which variables are worthwhile being observed. Finally, if a response is observed, we can trace the effect of the observation backward to identify particular influential causes. Therefore, the join-tree technique for GBFs can be used for influential findings as in non-monotonic reasoning.

In Example 4.1, we assume that each node in a join-tree has stored a GBF. Of course, this assumption can be removed by introducing vacuous belief functions. According to Section 2.1, a vacuous belief function is a special GBF with $\mathbf{B}^* = \mathbf{C}^* = \mathbf{V}^*$. Therefore, the combination and the marginalization for both GBFs and vacuous belief functions follow the Shafer–Shenoy axioms. i.e., the local computation scheme applies to a join-tree, in which some nodes may store vacuous belief functions. In Example 4.1, suppose we want to find the marginal of Bel on variable $Y$. We can simply attach a node, say $C_6$, which contains $Y$ only, to node $C_1$, and store a vacuous belief function for $Y$, denoted by $\phi_Y$, in $C_6$. After propagation, we will have $\mathrm{Bel}^{\downarrow Y} = \phi_Y \otimes m_{1\to 6} = m_{1\to 6}$.

## 5. Conclusion

In expert systems, the number of GBFs to be combined could be very large. It is inefficient and even infeasible to combine all of them first and then make inferences. The local computation scheme is suggested by Dempster [5] to solve this problem. However, because GBFs are neither finite nor condensable, Shafer [21] believes that Dempster's suggestion needs further justifications. One of the purposes of this work is to answer Shafer's call and to prove the feasibility of the local computation scheme for GBFs. This requires us, according to [28], to prove that the combination and marginalization of GBFs follow the Shafer–Shenoy axioms, which are the conditions for the local computation of any objects to be possible.

The primary difficulty in proving the Shafer–Shenoy axioms lies at interplaying combination and marginalization. As for normal distributions, marginalization can be easily defined in a variable space while combination (multiplication) in a sample space. To interplay combination and marginalization, we need define either marginalization in a sample space or combination in a variable space. In this work we take the latter approach. The combination rule in a variable space was initially proposed in [14]. However, because of its complicated and implicit representation, the rule is hardly useful for proving the Shafer–Shenoy axioms. Therefore, in this paper, we propose an alternative using matrix sweeping operators. The new rule reveals a concise and explicit link between the combined GBF and its components. Along with the properties of matrix sweepings, it underlies the third axiom of Shenoy and Shafer [28] in a tractable fashion. Also, as our examples illustrate, the new rule reduces the combination of GBFs to purely algebraic operations that can be implemented easily by spreadsheet programs.

To illustrate the feasibility of join-tree computation, we adopt the Shafer–Shenoy architecture of propagation and propose a local computation scheme for GBFs. We show briefly how to arrange all GBFs to be combined into a hypergraph, how to convert the hypergraph into a join-tree, and how to load the GBFs into the join-tree. We introduce two basic propagation rules and overall control strategies in sending and absorbing messages in a join-tree. We show how to compute messages in terms of GBFs and how to compute marginals for subsets of variables. We also present a comprehensive example to illustrate the proposed computation scheme and its potential usage in artificial intelligence and statistics.

The current work complements existing work on the computation of finite belief functions [9,24,28] and enriches the theory of local computation [23,20] by extending its applicability to a wide variety of problem domains such as knowledge integration, model combination, the Kalman filter, and empirical modeling. It also generalizes the existing work on the local computation of Gaussian probability distributions and of solving systems of linear equations.

## Acknowledgements

This paper benefited from many discussions with Glenn Shafer and the foundational work by Dempster [5] and Shafer [21]. The author also thanks Paul-Andre Monney and Amy Dolcourt for their constructive comments.

## References

[1] S. Arnborg, D.G. Corneil, A. Proskurowski, Complexity of finding embeddings in a $k$-tree, SIAM Journal of Algebraic and Discrete Mathematics 8 (1987) 277–284.
[2] G.E.P. Box, G.C. Tiao, Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, MA, 1973.
[3] A.P. Dempster, Elements of Continuous Multivariate Analysis, Addison-Wesley, Reading, MA, 1969.
[4] A.P. Dempster, Construction and local computation aspects of network belief functions, in: R.M. Oliver, J.Q. Smith (Eds.), Influence Diagrams, Belief Nets, and Decision Analysis, Wiley, Chichester, 1990.

[5] A. P. Dempster, Normal belief functions and the Kalman filter, Technical report, Department of Statistics, Harvard University, Cambridge, MA, 1990.
[6] J.E. Hunter, F.L. Schmidt, Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, Sage, Newbury Park, CA, 1990.
[7] F. Jensen, F.V. Jensen, S.L. Dittmer, From influence diagrams to junction trees, In: R.L. Mantaras, D. Poole (Eds.), Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence, San Francisco, CA, Morgan Kaufmann, Los Altos, 1994.
[8] F.V. Jensen, S.L. Lauritzen, K.G. Olesen, Bayesian updating in causal probabilistic networks by local computations, Computational Statistics Quarterly 4 (1990) 269–282.
[9] A. Kong, Multivariate belief functions and graphical models, Ph.D. thesis, Department of Statistics, Harvard University, Cambridge, MA, 1986.
[10] S.L. Lauritzen, D.J. Spiegelhalter, Local computation with probabilities on graphical structures and their application to expert systems (with discussion), Journal of the Royal Statistical Society Series B 50 (1988) 157–224.
[11] S.L. Lauritzen, N. Wermuth, Graphical models for associations between variables, some of which are qualitative and some quantitative, The Annals of Statistics 17 (1989) 31–57.
[12] L. Liu, Model combination using Gaussian belief functions, Technical report, School of Business, University of Kansas, Lawrence, KS, 1995.
[13] L. Liu, Propagation of Gaussian belief functions, in: D. Fisher, Hans-J. Lenz (Eds.), Learning Models from Data: AI and Statistics V, Springer, New York, NY, 1996, pp. 79–88.
[14] L. Liu, A theory of Gaussian belief functions, International Journal of Approximate Reasoning 14 (1996) 95–126.
[15] D. Maier, The Theory of Relational Databases, Computer Science Press, Rockville, MD, 1983.
[16] K. Mellouli, On the propagation of beliefs in networks using the Dempster–Shafer theory of evidence, Ph.D. thesis, School of Business, University of Kansas, Lawrence, KS, 1987.
[17] A.N. Michel, C.J. Herget, Applied Algebra and Functional Analysis, Dover, New York, NY, 1981.
[18] R.D. Shachter, C.B. Kenley, Gaussian influence diagrams, Management Science 35 (1989) 527–550.
[19] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[20] G. Shafer, An aximatic study of computation in hypertrees, Technical report, School of Business, University of Kansas, Lawrence, KS, 1991.
[21] G. Shafer, A note on Dempster's Gaussian belief functions, Technical report, School of Business, University of Kansas, Lawrence, KS, 1992.
[22] G. Shafer, Probabilistic Expert Systems, SIAM, Philadelphia, PA, 1996.
[23] G. Shafer, P.P. Shenoy, Local computation in hypertrees, Technical report, School of Business, University of Kansas, Lawrence, KS, 1988.
[24] G. Shafer, P.P. Shenoy, K. Mellouli, Propagating belief functions in qualitative Markov trees, International Journal of Approximate Reasoning 3 (1987) 383–411.
[25] P.P. Shenoy, A valuation-based systems for Bayesian decision analysis, Operations Research 40 463–484.
[26] P.P. Shenoy, A valuation-based language for expert systems, International Journal of Approximate Reasoning 3 (1989) 383–411.
[27] P.P. Shenoy, G. Shafer, Propagation belief functions using local computations, IEEE Expert 1 (1986) 43–52.
[28] P.P. Shenoy, G. Shafer, Axioms for probability and belief-function propagation, in: R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, vol. 4, North-Holland, Amsterdam, 1990, pp. 169–198.
[29] R.E. Tarjan, M. Yannakakis, Simple linear time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acylic hypergraphs, SIAM Journal of Computing 13 (1984) 566–579.

[30] F. Yates, An apparent inconsistency arising from tests of significance based on fiducial distribution of unknown parameters, Proceedings of the Cambridge Philosophical Society 35 (1939) 579.

[31] L. Zhang, Studies on finding hypertree covers for hypergraphs. Technical report, School of Business, University of Kansas, Lawrence, KS.